ROBUSTNESS AND POWER OF SOME ALTERNATIVE METHODS TO THE INDEPENDENT SAMPLES t-TEST WHEN A DIFFERENCE IN MEANS IS ACCOMPANIED BY A DIFFERENCE IN VARIANCES

By

MBULAHENI E. NTHANGENI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

ACKNOWLEDGMENTS

I am indebted to several individuals who assisted me in this project and throughout my doctoral program. First, I would like to thank Dr. James J. Algina, my dissertation chairperson, for suggesting the topic, guiding me through the whole process and providing invaluable support. Second, I would like to thank my committee members: Dr. Linda M. Crocker, for the continuous support; Dr. M. David Miller, for the financial support through Research Assistantship and for believing in me; and Dr. Sencer Yeralan, for the many years of support, from my master's program to this project. I also want to thank Dr. John M. Newell who provided me with encouragement throughout the project. Third, I would like to thank Dr. Jon Morris, from the Department of Advertising, for offering me a Research Assistantship when I had lost all hope. Fourth, I would like to thank my family: my wife, Gladys, for being so understanding and supportive during the hard times, and my children, Tshilidzi and Mpho, for being a source of my motivation. Finally, I would like to thank God without whom all this could not have been possible. It is a miracle indeed!

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

ABSTRACT iv

CHAPTER 1 INTRODUCTION AND LITERATURE REVIEW 1
Review of Literature 4
Statement of the Problem 22
Significance of the Study 24

CHAPTER 2 METHODOLOGY 26
Design of Study 26
Simulation Procedures 30
Summary 32

CHAPTER 3 RESULTS AND DISCUSSION 33
Estimated Type I Error Rates 35
Power Analyses 50
Power for Variance Differences Only 50
Power for Mean and Variance Differences . . . 67
Summary 79

CHAPTER 4 CONCLUSIONS 84
Conclusions 84
Theoretical Explanations of Selected Results . . . 86
Comparisons to Previous Results 88
Suggestions for Future Research 90

APPENDIX A ESTIMATED TYPE I ERROR RATES AND POWER . . . 94

APPENDIX B VARIANCE COMPONENTS FOR ESTIMATED TYPE I ERROR RATE AND POWER 112

REFERENCES 119

BIOGRAPHICAL SKETCH 122

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ROBUSTNESS AND POWER OF SOME ALTERNATIVE METHODS TO THE INDEPENDENT SAMPLES t-TEST WHEN A DIFFERENCE IN MEANS IS ACCOMPANIED BY A DIFFERENCE IN VARIANCES

By

Mbulaheni E. Nthangeni

December, 1995

Chairperson: James J. Algina
Major Department: Foundations of Education

It is common in educational and behavioral studies to encounter situations in which a difference in means is accompanied by a difference in variances. In this case, the independent samples t-test has been shown to work poorly. Several alternative methods have been suggested in the biometrics literature.

Three hundred and sixty simulated experiments were conducted to compare Type I error rates and power of the independent samples t-test, Welch's test, and the four alternative methods from the biometrics literature. The factors were (a) total sample size (40, 80, 120), (b) sample size ratio (2/3, 1, 3/2), (c) the ratio of standard deviations for the experimental and control groups (1/3, 1/2, 1, 2, 3) (d) power of the tests under normality and homoscedasticity (0, .5, .7, .9), and (e) type of distribution (normal,

lognormal).

Results of the study show that, with one exception, all of the methods investigated had good control of the Type I error rates. When data were lognormal, the test developed by Brownie, Boos, and Hughes-Oliver did not control the Type I error rate. The alternative methods had greater power than the independent samples t-test and Welch's test when the standard deviation ratio was larger than 1. Also, the alternative methods showed better power when data were sampled from the normal distribution than when the data was from the lognormal distribution. Of the four alternatives, a test suggested by Conover and Salsburg had the best power and appears to be the most promising for use in educational research.

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

A common belief among educational researchers, in particular, and educators, in general, is that the effectiveness of instructional methods depends on the type of students with whom the methods are used. Consider a situation in which the effectiveness of two instructional methods is measured by a criterion variable such as achievement. As depicted in Figure 1, the relative effectiveness of the two methods depends on the pretest level of the student, where the pretest variable is some predictor of the criterion. Each line in Figure 1 represents the regression of the criterion on the predictor for the students exposed to one of the two instructional methods. In this figure, the difference in the effectiveness is larger for the students who are high on the predictor than it is for students who are low on the predictor.

Consider an experiment comparing two instructional methods and suppose the researcher has hypotheses about the specific predictors on which relative effectiveness depends. The hypotheses can be tested by collecting data on the predictors and, for each predictor, testing equality of the slope across instructional methods. Now suppose the

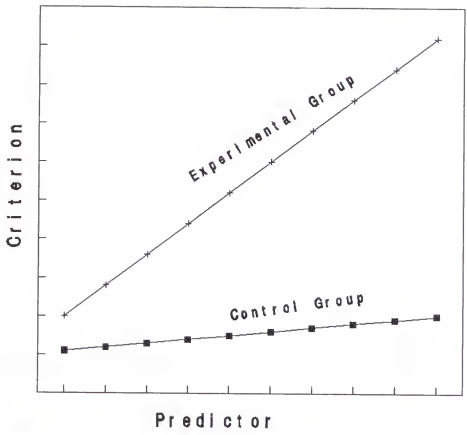

Figure 1: Treatment comparison in which treatment effects depend on predictor score.

researcher does not have any specific hypothesis or cannot narrow the number of possible predictors to a manageable number. How can the researcher proceed?

One procedure is to analyze the criterion data using the independent samples t-test. However, if the situation conforms, for example, to Figure 1 and if the conditional variances are equal in the two groups, the marginal variances will be unequal. Therefore the equality of variance assumption of the independent samples t-test will be violated. In Figure 1, H_o: $\mu_e - \mu_c = 0$ is false. Consequently, the researcher need not be concerned about the impact of the violation on the Type I error rate. Rather, the researcher should be concerned about whether other procedures are more powerful than the independent samples t-test when a difference in means is accompanied by a difference in variances.

In the biometrics literature (Brownie, Boos, & Hughes-Oliver, 1990; Conover & Salsburg, 1988; and Good, 1979), several procedures have been proposed for situations in which an experimental treatment is compared to a control and the mean or the mean and variance are expected to increase under the experimental treatment. An advantage of these procedures is that they have better power than the independent samples t-test when the mean and variance increase under the experimental treatment (Conover & Salsburg, 1988; and Good, 1979). The purpose of this

dissertation is therefore to compare the power of several procedures in situations like that depicted in Figure 1. The procedures selected for comparison are the independent samples t-test for means, Welch's test, and several of the alternative procedures from the biometrics literature.

Review of Literature

Assume that $Y_1, Y_2, \ldots, Y_n$ are independently and identically distributed with distribution function G_1 and $Y_{n+1}, Y_{n+2}, \ldots, Y_{m+n}$ are independently and identically distributed with distribution function G_2. Good (1979) investigated the case in which not all experimental subjects will respond to treatment and was interested in testing H_0: $G_2(Y) = G_1(Y)$ against the contaminated alternative H_1: $G_2(Y) = pG_1(Y-\Delta)+(1-p)G_1(Y)$, where p is the proportion of subjects who respond to the experimental treatment. Let μ_c be the center of G_1, the distribution from which control subjects are sampled. The alternative hypothesis states that 100(1-p)% of the experimental subjects come from a distribution identical to G_1, and 100p% come from a distribution that has center $\mu_c-\Delta$ but otherwise is identical to G_1. When all subjects are capable of responding (p=1.0), the uniformly most powerful test under normality is the independent samples t-test (Best & Rayner, 1987). A mixture of non-responders reduces the power of this test (Subrahmaniam, Subrahmaniam & Messeri, 1975). Hodges and Lehmann (1956)

found that the Wilcoxon test fares even worse. The reduction of power is due to a decrease in the absolute difference between means and to an increase in the variance of the treatment sample. As Good noted, this suggests that one consider test statistics that are directly proportional to both the difference in sample means and the difference in sample variance.

Good (1979) suggested the following statistic:

$$v(\theta) = \theta\left(\frac{1}{n_c}+\frac{1}{n_e}\right)^{-1}(\bar{Y}_c-\bar{Y}_e)^2+(1-\theta)SS_c \qquad (1)$$

where $\bar{Y}_c$, SS_c, n_c ($\bar{Y}_e$, SS_e, n_e) denote the mean, sum of squared deviations about the mean, and size of the control (treatment) sample and $0<\theta<1$. The test is conducted by using Fisher's method of randomization and is defined as follows: For a two tailed test, reject H_0 if $v(\theta)$ based on the original observations exceeds a proportion α of all computed values. For a one tailed test, reject H_0 if the treatment mean exceeds the control mean and if $v(\theta)$ based on the original observations exceeds a proportion 2α of all computed values.

The value that must be selected for θ depends on p and G_1. For example, if p=1.0 (i.e. all treatment subjects respond to treatment) and G_1 is normal, selecting θ=1.0 will maximize power. Setting θ=1.0 yields a test equivalent to Fisher's exact test. If G_1 is normal and p<1.0, setting

θ=p tends to maximize power (Good, 1979). If G_1 is lognormal, power tends to be maximized for θ<1.0 even when p=1.0 (Good, 1979). Of course p is unknown and precise calibration of θ to G_1 is unlikely. As a result Good suggested using θ=0.67.

Good (1979) conducted a small simulation and showed that when G_1 is normal, his test is more powerful than the independent samples t-test, provided p<1.0. Good included only conditions with equal sample sizes ($4 \leq n \leq 12$). According to Good, the test has a particular advantage when the sample size is small and Δ is large. However, inspection of Good's result suggests the advantage increases from n=4 to n=12. Good's simulation also showed that when G_1 is lognormal, his test is more powerful than the independent samples t-test ($.2 \leq p \leq 1.0$). However, it also has a slightly elevated Type I error rate when G_1 is lognormal. Good's conditions with G_1 lognormal were limited to $n_1 = n_2 = 4$.

The randomization procedure used in Good's procedure has a practical limitation. For small sample sizes the procedure can be used without any problems. However, for large sample sizes, the randomization procedure is computationally intensive, especially in simulation studies. Thus Good's test will not be included in the current study.

Contrary to Good's (1979) results, a study by Boos and Brownie (1986) showed that the Wilcoxon test was quite useful when $p \geq .6$, and they also pointed out an error in

Good's (1979) definition of the one-tailed randomization test. Boos and Brownie noted that Good's method for performing a one-tailed randomization test was not valid in general. Their study showed that for $p \geq .6$, the Wilcoxon test compares favorably with the independent samples t-test and with Good's test.

Mielke and Berry (1994) derived a permutation procedure for testing differences in location and scale among treatments in a completely randomized design. Consider a completely randomized experiment where $\Omega=\{\omega_1, \ldots, \omega_N\}$ denotes a finite sample of N subjects obtained from some super population. The sample is partitioned into g disjoint groups $S_1, \ldots, S_g$. Let

$$\Delta_{I,J}=|x_I-x_J|^v \qquad (2)$$

be the distance between the univariate response measurements of subjects ω_I and ω_J, where x_I denotes a response measurement for ω_I. If $v=1$, then $\Delta_{I,J}$ is the Euclidean distance between response measurements and the following is a test of between-group differences in location and/or scale. Let

$$\xi_k=\binom{n_k}{2}^{-1}\sum_{I,J}\Delta_{I,J}\psi_k(\omega_I)\,\psi_k(\omega_J) \qquad (3)$$

represent the average between-subject difference for all subjects within treatment group S_k (k=1,..,g), where the summation is such that $1 \le I \le J \le N$ and $\Psi_k(\omega_J)=1$, if $\omega_J \in S_k$. The average within-group difference weighted by the number of subjects n_k in treatment group S_k (k=1,..,g) is defined by

$$\delta = \sum_{k=1}^{g} C_k \xi_k \tag{4}$$

where $\sum C_k=1$, $C_k=n_k/L$ and $L=\sum n_k = N$.

The null hypothesis H_0 is that the sample values are independent of group membership. The probability value associated with an observed value of δ, δ_0, is the probability under H_0 of observing a value of δ as extreme or more extreme than δ_0. An exact probability value is expressed as

$$p=P[\delta \le \delta_0] = \frac{number\ of\ \delta's \le \delta_0}{M} \tag{5}$$

where

$$M = \frac{N!}{\prod_{k=1}^{g} n_k!} . \tag{6}$$

To compute the exact probability values, Mielke and Berry used a Pearson Type III distribution probability value approximation. Under H_o, the mean, variance and skewness of δ are given by

$$\mu_\delta = M^{-1}\sum_{j=1}^{M}\delta_j \ , \tag{7}$$

$$\sigma_\delta^2 = M^{-1}\sum_{j=1}^{M}\delta_j^2 - \mu_\delta^2 \ , \tag{8}$$

and

$$\gamma_\delta = \frac{M^{-1}\sum_{j=1}^{M}\delta_j^3 - 3\mu_\delta\sigma_\delta^2 - \mu_\delta^3}{\sigma_\delta^3}, \tag{9}$$

respectively. The standardized statistics of δ given by $T=(\delta-\mu_\delta)/\sigma_\delta$ is approximated, under H_o, by the Pearson Type III distribution with the density function given by

$$f(y) = \frac{(-\frac{2}{\gamma_\delta})^{\frac{4}{\gamma_\delta^2}}}{\Gamma(\frac{4}{\gamma_\delta^2})}\,[\frac{-(2+y\gamma_\delta)}{\gamma_\delta}]^{\frac{(4-\gamma_\delta^2)}{\gamma_\delta^2}}\ x\ \exp[\frac{-2(2+y\gamma_\delta)}{\gamma_\delta^2}] \tag{10}$$

where $-\infty < y < -2/\gamma_\delta$ and $\gamma_\delta < 0$. If $\gamma_\delta \geq 0$, then the density function of the Pearson Type III distribution is given by

$$f(y) = \frac{(\frac{2}{\gamma_\delta})^{\frac{4}{\gamma_\delta^2}}}{\Gamma(\frac{4}{\gamma_\delta^2})} [\frac{(2+y\gamma_\delta)}{\gamma_\delta}]^{\frac{(4-\gamma_\delta^2)}{\gamma_\delta^2}} \; x \exp[\frac{-2(2+y\gamma_\delta)}{\gamma_\delta^2}] \qquad (11)$$

where $-2/\gamma_\delta < y < \infty$ $-2/\gamma_\delta$ (Mielke, 1991). If $\gamma_\delta = 0$, $f(y)$ is a standard normal distribution given by

$$f(y) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{y^2}{2}). \qquad (12)$$

The probability value for an observed value of δ, δ_0, and, correspondingly, $T_0 = (\delta_0 - \mu_\delta)/\sigma_\delta$, is given by

$$p = P[\delta \leq \delta_0] = \int_{-\infty}^{T_0} f(y)\, dy. \qquad (13)$$

Mielke and Berry also derived a multivariate extension in which r response measurements are associated with subject ω_I (I=1,...,N). The distance function defined in (1) is now given by

$$\Delta_{I,J} = \left[\sum_{j=1}^{r} (x_{jI} - x_{jJ})^2\right]^{\frac{v}{2}} \qquad (14)$$

When $v=1$, $\Delta_{I,J}$ is an r-dimensional Euclidean distance between the response measurements. When there are more than one response measurements for each object, if the response measurement are expressed in different units they must be made to commensurate, that is, standardization among multivariate responses must be made. For example, if $y'_I=(y_{1I},\ldots,y_{rI})$ for $I=1,\ldots,N$ denote N noncommensurate r-dimensional values, the corresponding N commensurate r-dimensional values denoted by $x'_I=(x_{1I},\ldots,x_{rI})$ are given by $x_{jI}=y_{1I}/\phi_j$ where

$$\phi_j=\left[\sum_{I<J}|y_{jI}-y_{jJ}|^v\right]^{\frac{1}{v}} \tag{15}$$

for $j=1,\ldots,r$. The commensurated data have the property that

$$\sum_{I<J}|x_{jI}-x_{jJ}|^v=1 \tag{16}$$

for $j=1,\ldots,r$ and any choice of v. This commensuration procedure is called Euclidean commensuration when $v=1$ (Berry & Mielke, 1992). An alternative commensuration procedure called Hotelling commensuration involves the distance

$$\Delta_{I,J}=\left[(x_I-x_J)'S^{-1}(x_I-x_J)\right]^{\frac{v}{2}} \tag{17}$$

where S is the r x r variance-covariance matrix.

The procedure for computing δ, M, and p based on the Pearson Type III distribution for the multivariate case is the same as the univariate case. When the Hotelling commensuration is used, with $g \geq 2$ and $\upsilon=2$, δ is related to the Bartlett-Nanda-Pillai trace statistic for multivariate analysis of variance as follows:

$$\delta = \frac{2(r-V)}{N-g} . \qquad (18)$$

Mielke and Berry (1994) conducted Monte Carlo simulations to obtain power comparisons for multivariate permutations tests. The power comparisons of the bivariate data included five distributions: bivariate normal, bivariate uniform, bivariate exponential, bivariate lognormal, and bivariate Cauchy. The comparisons were made at $\alpha=0.01$ and $\alpha=0.05$. Each comparison involved two groups, where the first was the control group and the second was the experimental group differing from the control group by a shift in location. The location shift for the normal, uniform, exponential, lognormal and Cauchy were, respectively, 0.54, 0.20, 0.47, 0.73, and 1.70. The sample sizes considered were $n_c=n_e=20$. The three permutation tests consisted of (a) Euclidean commensuration, (b) Hotelling commensuration with $\upsilon=1$ and Hotelling commensuration with

υ=2. Also included was the Bartlett-Nanda-Pillai test. Power results showed that the Euclidean commensuration permutation test performed quite well for the five distributions considered. The Hotelling commensuration with υ=2 and Bartlett-Nanda-Pillai test performed very well for both the bivariate normal and bivariate uniform. The Euclidean commensuration test performed better than the other tests for the bivariate normal at α=0.01. For the bivariate exponential, bivariate lognormal, and bivariate Cauchy distributions, the Euclidean commensuration permutation test and Hotelling commensuration test with υ=1 performed better than Hotelling commensuration test based on υ=2. The Euclidean commensuration permutation test performed better than the Hotelling commensuration test with υ=1 for all five distributions. The actual Type I error rate for the Bartlett-Nanda-Pillai test were substantially lower than the nominal Type I error rate for the bivariate exponential, bivariate lognormal, and bivariate Cauchy distributions.

Johnson, Verrill and Moore (1987) derived a locally most powerful rank test H_0 against

$$H_2: G_2(Y)=pG_3(Y)+(1-p)G_1(Y)$$

where $G_3(Y)\leq G_1(Y)$ for all Y. Johnson et al. studied two special cases of H_2. In the first case G_1 and G_3 are normal cumulative density functions with common σ and $G_3(Y)=G_1(Y-\Delta)$, so H_2 specializes to H_1. In the second $g_1(Y)$ and $g_3(Y)$, where g_1 and g_3 are densities corresponding to G_1 and G_3, are

bi-uniform densities. For each the range of Y is (a,d] and the step in the density occur at Y=c. The locally most powerful test for H_0 and H_2 rejects H_0 for large values of

$$\sum_{i=1}^{n} E\left[\frac{g_3(V^{(r_i)})}{g_1(V^{(r_i)})}-1\right] \qquad (19)$$

where $V^{(1)}\leq V^{(2)}\leq.....\leq V^{(m+n)}$ is an ordered sample from $G_1(x)$ and $r_1\leq r_2\leq.....\leq r_n$ are the ranks of $Y_1,....,Y_n$ in the combined sample. In Equation 19, g_1 and g_3 are the densities of $G_1(x)$ and $G_3(x)$ respectively. An approximation of the locally most powerful rank test is given by replacing the expectation in Equation 19 by the score

$$J\left(\frac{r_i}{m+n+1}\right) \qquad (20)$$

where J(u), the score function, is defined on the interval $0<u<1$ by

$$J(u)=\frac{g_3[G_1^{-1}(u)]}{g_1[G_1^{-1}(u)]}-1 \qquad (21)$$

For large samples sizes, Johnson et al. noted that there was no difference between the locally most powerful rank test and the approximate test.

When the alternative hypothesis is H_1, a mixture of normals (MN), the score function is

$$J_{MN}(u) = e^{-\frac{\Delta^2}{2\sigma^2}} e^{\frac{\Delta}{\sigma}\Phi^{-1}(u)} - 1 \qquad (22)$$

where Φ is the standard normal cumulative distribution function. Note that to use the score function Δ must be specified. When G_1 and G_3 are bi-uniform (BU) distributions the score function is

$$J_{BU}(u) = \begin{cases} 0, & 0 < u \leq b_1 \\ 1, & b_1 < u < 1 \end{cases} \qquad (23)$$

where

$$g_1(Y) = \begin{cases} \dfrac{b_1}{(c-a)}, & a < Y \leq c \\ \dfrac{(1-b_1)}{(d-c)}, & c < Y \leq d \end{cases} \qquad (24)$$

and

$$g_3(Y) = \begin{cases} \dfrac{b_2}{(c-a)}, & a < Y \leq c \\ \dfrac{(1-b_2)}{(d-c)}, & c < Y \leq d \end{cases} \qquad (25)$$

To use the score function the parameter b_1 in the density g_1 must be specified.

By using the asymptotic relative efficiencies and simulation results, Johnson et al. showed the following:

1. When J_{MN} is the correct score function, the relative efficiency of the Wilcoxon or normal scores test is <1.0. Similarly, when J_{BU} is the correct score function the

relative efficiency of the Wilcoxon or normal scores test is <1.0.

2. When J_{BU} is the correct score function the relative efficiency of the test using J_{MN} is <1.0.

3. When J_{MN} is the correct score function the relative efficiency of J_{MN} using a misspecified Δ is approximately 0.8. When the true and specified Δ/σ are different by 0.5, when the true difference is 1.0, relative efficiency is about 0.5.

4. When J_{MN} is the correct score function, use of J_{BU} can result in Type I error rates that are substantially too high, particularly when sample sizes are small.

5. When J_{MN} is the correct score function, use of Wilcoxon or normal scores test results in a loss in power that increases as sample sizes increases and increases dramatically as Δ increases.

6. When J_{MN} is the correct score function, Good's (1979) test has power comparable to the test using J_{MN}.

Conover and Salsburg (1988) noted that Good (1979)'s alternative hypothesis is a special case of

$$H_2: G_2(Y) = pG_3(Y) + (1-p)G_1(Y) \qquad (26)$$

where $G_3(Y) \neq G_1(Y)$. In Good's paper $G_3(Y) = G_1(Y-\Delta)$. Conover and Salsburg suggested two other alternative hypotheses based on the Lehmann alternative. The first is

$$H_3: G_2(Y) = p[G_1(Y)]^{\tau} + (1-p)G_1(Y) \qquad (27)$$

where $\tau \geq 1$. This alternative implies that measurement of a subject who responds to the treatment has the same distribution as the maximum of τ responses by controls. Under H_3 the range of scores for the experimental and control subjects is the same. The test statistic is based on the locally most powerful rank test. The alternative hypothesis H_3 has two parameters p and τ. Consideration of each parameter by itself leads to two different cases, each with its own locally most powerful rank test. To test the H_0 against H_3 as p approaches 0^+, the locally most powerful rank test uses the statistic

$$T = \sum_{i=1}^{N} s_1(R_i) \qquad (28)$$

where $s_1(i) = (N)_{N+1-i}/(N+\tau-1)_{N+1-i}$ are the scores for the locally most powerful rank test. In the expression for s_1, $(N)_k = N(N-1).....(N-k+1)$. Conover and Salsburg point out that an alternative to s_1 is

$$s_2(i) = \left(\frac{i}{N+1}\right)^{\tau-1} \tag{29}$$

Both tests have the same asymptotic relative efficiency (ARE), but $s_2(i)$ is easier to use. The parameter τ is unknown. Conover and Salsburg suggest using $\tau=5$ and note that $\tau=2$ yield the Wilcoxon rank sum test (Lehman, 1953). Conover and Salsburg point out that use of Savage scores

$$s(i) = -\sum_{j=i}^{N} \frac{1}{j} \tag{30}$$

yields the locally most powerful test for τ. They report that the test has good power when $\tau=1$, but not when $3 \leq \tau \leq 5$ which they argue is common in biometrics data.

Conover and Salsburg's second alternative hypothesis is

$$H_4: \; G_2(Y) = p\left[1-[1-G_1(Y)]^{\frac{1}{\tau}}\right] + (1-p)\,G_1(Y) \tag{31}$$

which implies that the response distribution for control subjects is the same as the distribution of the minimum of τ experimental responses. The test for the second alternative hypothesis is similar to that for the first alternative, with the scores for the locally most powerful test with τ approaching 1^+, given by

$$s_3(i) = \sum_{j=1}^{i} \frac{1}{N+1-i} \qquad (32)$$

Scores that have the same ARE as s_3 but are easier to calculate are given by

$$s_4(i) = -\ln\left[1 - \frac{i}{N+1}\right] \qquad (33)$$

Conover and Salsburg do not report a test for parameter p in the second alternative hypothesis. Critical values for the methods can be obtained for small samples by tabulating the randomization distribution of T. For large samples Conover and Salsburg suggest applying the independent samples t-test to the scores. They report that a simulation study supports this method even when there are as few as five subjects in each group.

Conover and Salsburg (1988) applied the methods,using s_1, s_2, s_3 and s_4, the method due to Johnson, Verrill, and Moore, Good's test, and the independent samples t-test to two data sets. They noted that (a) use of s_1 and s_2 yields the smallest p values and (b) the p values for s_1 and s_2, and s_3 and s_4 are fairly similar. Recall that s_2 is an approximation to s_1 and s_4 is an approximation to s_3.

Brownie, Boos and Hughes-Oliver (1990) proposed a modification of the independent samples t-test and F-test to achieve better power in situations where increased variability is expected to be a consequence of treatment

effect. To test the hypothesis

$$H_0 : \mu_c - \mu_e, \sigma_c^2 - \sigma_e^2 \quad vs \quad H_a : \mu_c \neq \mu_e, \sigma_e^2 - \rho\sigma_c^2 , \ 1 \leq \rho \leq \infty \qquad (34)$$

the proposed modification of the independent samples t-test is given by

$$t_m = \frac{\overline{Y}_e - \overline{Y}_c}{[S^2_c(\frac{1}{n_c} + \frac{1}{n_e})]^{\frac{1}{2}}} \qquad (35)$$

with n_c-1 degrees of freedom. For testing equality of three or more means, the modified F is given by

$$F_m = \frac{\sum_{i=1}^{k} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2/(k-1)}{s_c^2} \qquad (36)$$

with k-1 and n_c-1 degrees of freedom.

Brownie et al. conducted simulations to compare the independent samples t-test and the modified t-test for given n_c, n_e, $\Delta_1 = (\mu_c - \mu_e)/\sigma_c$, $\rho = \sigma_c^2/\sigma_e^2$ and significance level α. The power curves showed that for fixed n_c, n_e, and Δ_1 the power gain for the modified t-test increases as ρ increases. The difference in power of the two procedures also depends on the sample size. If $\rho > 1$ then for fixed Δ and $n_c \geq n_e$, the power advantage of the modified t-test increases as the proportion of observation in the treatment group increases.

Brownie et al. (1990) warn that the decision to use the modified procedure should be made a priori, and not because

variability appears low for some subset of the groups. Also, the modified t-test and F-test should not be used when the number of control observations is small (n_c<10) because of the power calculations and because of a general stability of the variance estimate.

Blair and Sawilosky (1993) investigated the impact of violating the normality assumption on the Type I error rate and power of the modified F-test, F_m, suggested by Brownie et al.(1990). They included four distributions: uniform, t with df=3, exponential and chi-square with df=1. The number of groups varied from 2 to 10 in steps of 2 and sample size per cell was either 10, 12, 40 or 80. For the F_m tests, estimated actual Type I error rates were below nominal levels for the uniform distribution and above nominal levels for the remaining distributions. The discrepancy between nominal and estimated actual Type I error rates increased as the number of groups increased and declined as the sample size per cell increased. For two groups and α=0.05, the maximum discrepancy between nominal and estimated actual Type I error rates was -.003, .021,0.39, and .068 for the uniform, t(3), exponential and chi-square (1) distributions. As expected from theoretical and empirical results, the F test was generally quite robust. Power results showed that F_m was less powerful than F when the groups differed only in mean, with the power disadvantage increasing as the number of groups increased. The F_m test had larger power when

groups differed in mean and variance, with the power advantage increasing as the number of groups increased. Blair and Sawilosky concluded that the robustness of the fixed-effect ANOVA to departures from population normality does not carry over to the modification proposed by Brownie et al., contrary to the conclusion by Brownie et al.

Brownie et al. also pointed out that the Welch technique is designed for

$$H_0: \mu_c=\mu_e,\ \sigma_c^2\neq\sigma_e^2 \quad \text{vs } H_1: \mu_c\neq\mu_e,\ \sigma_e^2\neq\sigma_c^2 \qquad (37)$$

and not for

$$H_0: \mu_c=\mu_e,\ \sigma_c^2=\sigma_e^2 \quad \text{vs } H_1: \mu_c\neq\mu_e,\ \sigma_e^2=\rho\sigma_c^2. \qquad (38)$$

The Welch test measures the distance between the control and treatment means $(\bar{Y}_e - \bar{Y}_c)$ in terms of $(\sigma_c^2/n_c + \sigma_e^2/n_e)^{1/2}$. If $n_c=n_e$, the power of Welch's test cannot exceed the power of the independent samples t-test. For $n_e>n_c$ and $\rho>1$, the independent samples t-test is more powerful than Welch's test. Thus for $n_e\geq n_c$ and $\rho>1$, Welch's test is less powerful than the independent samples t-test and hence less powerful than the modified t-test. However, if $n_e<n_c$ and $\rho>1$ Welch's test is more powerful.

Statement of the Problem

In educational and psychological experimentation, researchers are likely to be confronted with situations in which differences in means are accompanied by differences in variance. In this situation, tests such as the independent

samples t-test, which assumes homogeneity of variance, may be less powerful than techniques designed for use when differences in means accompany differences in variance. Several such procedures have been developed in the biometrics literature.

The procedures developed in the biometrics literature fall into two classes. The first, including the techniques developed by Johnson et al. (1987) and Conover and Salsburg (1988), comprise tests that are the most powerful for certain alternative hypotheses. However, all the alternative hypotheses imply that differences in means will be accompanied by differences in variance. If the data do not conform to the appropriate alternative hypothesis, power will be negatively impacted. It is therefore of interest to study the power of these tests when differences in means are accompanied by differences in variance, but the data do not conform to the specific hypotheses for which these test were designed. The second class of tests, including those developed by Good (1979), Brownie et al. (1990), and Mielke and Berry (1994) comprises tests that (a) are designed to be more powerful than the independent samples t-test when differences in means are accompanied by differences in variances but (b) do not specify a particular distribution in the alternative hypothesis. These tests should be more powerful than the independent samples t-test in a situation like that depicted in Figure 1. Little is known about the

degree to which these test are more powerful because available analytic and simulation results cover such a small range of conditions.

The purpose of the present study was to compare Type I error rate and power of the alternative procedures to the independent samples t-test when differences in means are accompanied by differences in variances, by means of a Monte Carlo simulation technique. The procedures investigated are those developed by Johnson et al. (1987), Conover and Salsburg (1988), and Brownie et al.(1990). The procedure developed by Good (1979) and Mielke and Berry (1994) are not included because the permutations required for these tests make them expensive to investigate by simulation methods. The performance of these procedures were compared with the independent samples t-test and Welch's test. The procedures were investigated for both equal and unequal sample sizes.

Significance of the Study

A variety of procedures for comparing means when variance differences are expected have been presented in the literature. However, these procedures have not been compared with one another, so it is difficult to choose among the procedures. This study will provide results that should prove useful in making a choice between the independent samples t-test and one or more of the alternative procedures, and thereby enhancing the probability of

accurate and valid conclusions in educational experimentation.

CHAPTER 2
METHODOLOGY

In this chapter, the design of the study and the simulation procedures employed to carry out the study are described. The factors investigated in the study included total sample size, sample size ratio, mean differences, standard deviation ratio, and type of distribution. A Monte Carlo simulation technique was used to investigate the effect of the factors on the Type I error rate and power of the alternative procedures. The alternative procedures investigated were Conover and Salsburg's tests, Johnson, Verrill and Moore's test for mixed normal distributions, and Brownie, Boos and Hughes-Oliver's modified t-test. For the Johnson et al. test Δ/σ was set to 0.6. These procedures are designed to test hypotheses about differences between means of two populations when differences in means are accompanied by differences in variances. In addition, the independent samples t-test and Welch's test were also included for comparison with the alternative procedures.

Design of Study

The probabilities of Type I error and power were investigated in connection with the factors listed in the

preceding paragraph. The levels of these factors are presented in the following paragraphs.

Total sample size. Three total sample sizes were included in the study: N=40, N=80, N=120. These sample sizes span the range from small to fairly large studies.

Type of distribution. Two types of distributions were included in the study: the normal and the lognormal distributions. The choice of distributions leads us to a question of whether the data in the study is representative of what is found in practice, especially the lognormal distribution. Miceri (1989) conducted a study of skewness of 440 univariate distributions including 231 distributions of standardized achievement and ability test scores, 35 distributions of criterion-refenced and mastery test scores, 125 distributions of personality test scores, and 49 distributions of gain scores based on the preceding types of test scores. Miceri used skewness and two other criteria to classify distributions with respect to asymmetry. The cutoff scores on skewness were .32, .71 and 2.00. Of the 440 distributions 30.9% were classified at above .71. Of all the distributions 11.4% were classified at the category defined by 2.00 cut score. In this study, data generated from the lognormal distribution were adjusted so that skewness was 1.75. Thus, the data in the study are typical of what is seen in the literature.

Sample size ratio. In addition to conditions with equal sample sizes [(n_e, n_c) = (20,20); (40,40); and (60,60)], conditions with unequal sample sizes were employed [(n_e, n_c) = (24,16); (16,24); (48,32); (32,48); (72,48); and (48,72)]. Thus the sample size ratio was either 2/3, 1.0 or 1.5. Although many simulations employ a wider range of sample size ratios, a sample size ratio of 1.5 is fairly large in randomized experiments.

Mean difference. Four mean differences were employed in the study. One mean difference was zero. This permits estimation of Type I error rates, and power of the alternative procedure to detect variance differences. The other three mean differences were selected so that, when the variances were equal and α=0.05, the independent samples t-test would have power of 0.50, 0.70, or 0.90. These power levels were selected to cover a range of power from moderate to high. The effect sizes that result in these levels of power are shown in Table 1 for all sample size combinations. In Table 1, we assume that the mean for the control group is 0 and the common standard deviation is 1.

Standard deviation ratio. Five levels of standard deviation ratio σ_e/σ_c were employed : 3, 2, 1, 0.5 and 1/3. The ratios greater than 1 are conditions under which the alternative tests are expected to be more powerful than the independent samples t-test. The ratios less than 1 are included to investigate how well the alternative procedures

Table 1

Effect sizes (α=0.05)

Group sizes		Power		
n_c	n_e	.50	.70	.90
20	20	.53	.69	.94
24	16*	.54	.71	.96
40	40	.37	.49	.66
48	32*	.38	.50	.67
60	60	.30	.40	.54
72	48*	.31	.41	.55

* Interchanging n_e and n_c yields the same power when variances are homogeneous.

work in conditions other than those for which they were designed. When σ_e/σ_c = 3 or 2, σ_c=1. When σ_e/σ_c = 1/2 or 1/3, σ_e=1.

The choice of the standard deviation ratio is not unrepresentative of what occurs in practice. In his comments on choosing a statistics in multivariate analysis of variance, Stevens (1979) investigated subgroup variance differences of several real life data from major journals in educational and psychological research. He observed that even though subgroup variance differences as high as 36 times were rare, subgroup variance differences as high as 9 times were common. Also, Wilcox (1987) observed that among 14 studies found in the *American Educational Research Journal*, 3 reported sample variance ratios greater than 16. Wilcox argued that equal variances should not be taken for granted, when conducting educational research. Thus, Stevens (1979) and Wilcox (1987) results show that the choice of standard deviation ratio in this study is within the range of what is commonly found in practice.

Simulation Procedures

Random Number Generation Procedure

The simulation in this study was carried out using the random number generating functions in the Statistical Analysis System (1985). These random number generators are described below.

Normal Distribution. The normal random deviates were generated by employing the RANNOR generator in SAS-IML.

Lognormal Distribution. To generate the lognormal scores the generator RANNOR was used to generate the normal deviates (Z) and the lognormal scores were then obtained from the normal through transformation Y=exp(Z). In general, if Z is $N(\mu,\sigma)$ then Y is lognormal with mean

$$\mu_{LN} = \exp(\mu + \sigma^2/2)$$

and variance

$$\sigma_{LN}^2 = [\exp(2\mu + \sigma^2)][\exp(\sigma^2)-1]$$

(Mood, Graybill & Boes, 1974). In order to obtain an acceptable skewness, the transformation from normal to lognormal was adjusted so that the skewness for the lognormal distribution was 1.75. Setting μ=0 and σ=0.5, gives the required skewness.

Data Generation

The data were generated by using the following steps for each condition:

1. Generate n_c observations from the target population.

2. Generate n_e observations from the target population.

3. For the lognormal distributions, standardize the distributions using the population expected value and standard deviation.

4. Multiply each of the observations generated in step 2 by the appropriate constant to simulate

heteroscedasticity.

5. For non-null conditions add the appropriate constant to each observation in the experimental group.

For each replication, the data were analyzed by using the independent samples t-test, Welch's t-test and the alternative tests listed in the first paragraph of this chapter.

Summary

A total of 360 experimental conditions were simulated, with 5000 replications for each condition. These conditions can be classified into three groups:

1. There were 18 conditions with $\mu_e-\mu_c=0$ and $\sigma_e=\sigma_c$. These conditions were used to determine the Type I error rates.

2. There were 72 conditions with $\mu_e-\mu_c=0$ and $\sigma_e\neq\sigma_c$. These conditions were used to determine the power for detecting variance differences only.

3. There were 270 conditions for the conditions with $\mu_e-\mu_c>0$, $\sigma_e=\sigma_c$ and $\mu_e-\mu_c>0$, $\sigma_e\neq\sigma_c$. These conditions were used to determine the power to detect differences when means differ.

CHAPTER 3
RESULTS AND DISCUSSION

In this chapter analyses of Type I error rates and power are presented for tests conducted at α=0.05. The data used in these analyses are reported in Appendix A. Results in this chapter are reported in three sections: (a) estimated Type I error rates, (b) estimated power for variance differences only, and (c) estimated power for mean and variance differences. For each section, a summary of the estimated Type I error rates or power is presented. Then a summary of the analysis of variance (ANOVA) used to investigate the effects of the within-subjects factor and between-subjects factors on the Type I error rates or power is presented. For the Type I error rates analyses, the between-subjects factors were total sample size (N), sample size ratio (n_e/n_c), and distribution (D); and the within-subject factor was method (M). For the power analyses, two additional between-subject factors, standard deviation ratio (σ_e/σ_c) and target power (P), were included.

Because many of the factors that affect Type I error rates and power were included in the study, each ANOVA yielded a large number of significant effects. To determine which of the effects to interpret, variance components were

computed for all the effects. A mean square component, θ^2, for each effect was computed using the formula [(MSF-MSE)/NT] (Myers,1979), where MSF was the mean square error for the given effect, MSE was the between-subjects mean-square error when the given effect was between-subjects effects and within-subjects mean square error when the given effect was the within-subject effect or the interaction of the within-subject factor and between-subject factor(s), and NT was the number of levels for the factors not included in the given effect. Negative mean square error components were set to zero. For each effect with a non-negative mean square component, the proportion of variance component was computed using the formula θ^2/(SMSF+MSEB+MSEW), where SMSF is the sum of the mean square errors of all the effects, MSEB is the between-subject mean square error and MSEW is the within-subjects mean square error. The variance components values were used to select significant effects for interpretation. An effect was considered important if (a) it was significant and (b) the variance component was large relative to the variance component associated with other significant effects. To compromise between interpreting a large number of unimportant effects and failing to interpret an important effect, the criterion used to select effects for interpretation was the proportion of variance component greater than 1%. Other researchers have used this criterion (Coombs, 1992; and Chirembo, 1995).

Estimated Type I Error Rates

Table 2 shows a summary of the distribution of the estimated Type I error rates for the six methods investigated in the study. By Bradley's liberal(1978) criterion, a test is robust if $.5\alpha \leq \tau \leq 1.5\alpha$, where α is the nominal level of significance and τ is the true probability of Type I error rate. Otherwise, the test is either conservative or liberal. Using Bradley's liberal (1978) criterion, only the BT method is liberal in any conditions. However, the JT method appears to have slightly elevated Type I error rates. The TT and WT methods, and the two methods suggested by Conover and Salsburg (1988), CT1 and CT2, have estimated Type I error rates closest to the nominal level of significance. Tables 3 and 4 show a summary of the distributions of the estimated Type I error rates for the six methods when data were generated from the normal and lognormal distributions, respectively. Results in these tables indicate that when data are generated from the lognormal distribution, the BT method tends to have inflated Type I error rates, while the other methods maintain control of the Type I error rates. When data are generated from the normal distribution, all the methods appear to maintain control of the Type I error rates, by Bradley's criterion. However, when data are generated from either the normal and lognormal distributions, Type I error rates for the JT method are slightly elevated. Thus, when data are lognormal,

Table 2

Distributions of Estimated Type I Error Rates for the Six Methods (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.045	.045	.046	.050	.052	.056	.056
WT	.040	.043	.046	.049	.052	.058	.060
BT	.047	.047	.051	.061	.078	.096	.097
CT1	.045	.045	.047	.051	.052	.054	.054
CT2	.045	.046	.048	.049	.051	.055	.056
JT	.047	.048	.050	.053	.055	.060	.061

Note. TT = independent samples t-test; WT = Welch test;BT = Brownie et al.'s modified t-test; CT1 = Conover and Salsburg's s_2 test; CT2 = Conover and Salsburg's s_4 test; JT = Johnson et al.'s mixed normal test.

Table 3

Distributions of Estimated Type I Error Rates for the Six Methods: Normal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.046	.046	.048	.050	.051	.053	.053
WT	.046	.046	.048	.050	.051	.053	.053
BT	.047	.047	.047	.051	.052	.054	.054
CT1	.045	.045	.047	.051	.051	.054	.054
CT2	.047	.047	.049	.050	.051	.056	.056
JT	.049	.049	.052	.053	.054	.061	.061

Table 4

Distributions of Estimated Type I Error Rates for the Six Methods: Lognormal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.045	.045	.046	.050	.052	.056	.056
WT	.040	.040	.045	.046	.055	.060	.060
BT	.068	.068	.076	.078	.087	.097	.097
CT1	.045	.045	.047	.049	.052	.053	.054
CT2	.045	.045	.046	.049	.051	.055	.055
JT	.047	.047	.050	.051	.057	.060	.060

the BT method is not recommended.

Analysis of variance (ANOVA) was used to investigate the effects on the Type I error rates of the following factors: total sample size (N), sample size ratio (n_e/n_c), distribution (D), and method (M). All the factors were between-subject factors except for M, which was a within-subject factor. Because there were many significant main and interaction effects, only the significant effects that contributed to more than 1% of the total variance were considered for interpretation. These effects are presented in Table 5. The results for Type I error rates for tests conducted at α=0.01 and α=0.10 are similar to those reported in Table 5. Variance components for these tests are reported in Appendix B.

The M, D and MxD effects. Shown in Table 6 are mean values of the Type I error rates for the MxD interaction. Results in this table indicate that when data are generated from the lognormal distribution, the Type I error rate for the BT method is substantially larger than for the other methods. The other tests have Type I error rates near α. When data are generated from the normal distribution, all the methods have similar Type I error rates and maintain control of the Type I error rates. Similar observations were made earlier from the summary of the distributions of the Type I error rates for the normal and lognormal distributions. A theoretical explanation for the behavior of

Table 5

Percent of Total Variance for Estimated Type I Error Rates Effects Accounting ≥1% of the Variance (α=0.05)

Effect	Percentage
MxN	2.46
$Mxn_e/n_c xD$	3.85
MxNxD	4.41
Mxn_e/n_c	5.66
D	5.68
M	22.46
MxD	47.38

Table 6

Mean Estimated Type I Error Rates as a Function of Method and Distribution (α=0.05)

	Distribution	
Method	N	L
TT	.050	.050
WT	.050	.050
BT	.050	.081
CT1	.050	.050
CT2	.050	.052
JT	.053	.053

BT, when data is lognormal, is given in Chapter 4. In Table 7, the marginal mean values of the estimated Type I error rates for the six methods are presented. Results in this table show that the BT method has higher Type I error rates than the other methods. This is due to the distribution effect on the BT method, as shown in Table 6. The other methods appear to maintain control of the Type I error rates. Again, similar conclusions were made from the summary of the distribution of the Type I error rates. Table 8 shows the marginal means of the Type I error rates for the normal and lognormal distributions. When data are generated from the lognormal distribution the Type I error rates are slightly inflated; with data generated from the normal distribution Type I error rates very close to the nominal level of significance. The inflated results of the lognormal distribution are due to the effect of the BT method.

The MxN and MxNxD effects. In Table 9, the mean estimated Type I error rates for the MxNxD interaction are reported. As in previous tables, the results indicate that when data are lognormal, the Type I error rates for the BT method are elevated. The Type I error rate for BT declines with increasing N. For all the methods, when data are generated from the normal distribution, the total sample size does not appear to have an effect on the Type I error rates. In Table 10, mean values of the Type I error rates are presented for the MxN interaction. The table shows that

Table 7

Mean Estimated Type I Error Rates as a Function of Methods (α=0.05)

Method	Means
TT	.050
WT	.049
BT	.066
CT1	.050
CT2	.050
JT	.053

Table 8

Mean Estimated Type I Error Rates as a Function of Distribution (α=0.05)

Distribution	Means
N	0.051
L	0.055

Table 9

Mean Estimated Type I Error Rates as a Function of Method, Sample Size and Distribution (α=0.05)

		Size		
Distrb.	Method	40	80	120
N	TT	.051	.050	.048
	WT	.051	.051	.048
	BT	.051	.050	.049
	CT1	.051	.049	.050
	CT2	.051	.049	.051
	JT	.056	.052	.052
L	TT	.051	.049	.049
	WT	.050	.049	.049
	BT	.093	.079	.072
	CT1	.050	.050	.048
	CT2	.050	.050	.048
	JT	.056	.052	.050

Table 10

Mean Estimated Type I Error Rates as a Function of Method and Sample Size (α=0.05)

	Size		
Method	40	80	120
TT	.051	.050	.048
WT	.050	.050	.048
BT	.072	.064	.060
CT1	.051	.049	.049
CT2	.050	.050	.050
JT	.056	.052	.051

the Type I error rate for the BT method is above α but improves as the total sample size increases. This is primary because of the elevated Type I error rates for BT when the data are lognormal. For small total sample size, the JT method also tends to have a Type I error rate above the nominal level of significance.

The Mxn_e/n_c and Mxn_e/n_cxD effects. In Table 11, results of the estimated Type I error rates for the Mxn_e/n_cxD interaction are presented. The results indicate that, when data are generated from the lognormal distribution, Type I error rate for the WT and BT methods increases as n_e/n_c increases. For the other methods, Type I error rate declines as n_e/n_c increases. When data are generated from the normal distribution, the table shows that (a) the Type I error rates for the TT and WT methods do not appear to be affected by n_e/n_c, and (b) the Type I error rate for the BT method is smaller for n_e/n_c =0.67 than for n_e/n_c=1.5, while for the CT1, CT2, and JT methods the opposite is the case. Algina, Oshima, and Lin (1994) give a theoretical explanation for the behavior of TT and WT, when data are lognormal. For the BT method, a theoretical explanation is given in Chapter 4. Table 12 reports the means for the Mxn_e/n_c interaction. Results in the table show that for a balanced design (n_e/n_c=1), the BT method is the only method with substantially inflated Type I error rates. When n_e/n_c=0.67, both the BT and JT methods have inflated Type I

Table 11

Mean Estimated Type I Error Rates as a Function of Method and Sample Size Ratio and Distribution (α=0.05)

Distrb.	Method	n_e/n_c .67	1.0	1.5
N	TT	.049	.051	.050
	WT	.049	.051	.050
	BT	.047	.051	.051
	CT1	.053	.051	.046
	CT2	.052	.050	.049
	JT	.056	.053	.051
L	TT	.055	.047	.047
	WT	.043	.047	.058
	BT	.078	.080	.086
	CT1	.053	.048	.047
	CT2	.054	.048	.047
	JT	.057	.051	.050

Table 12

Mean Estimated Type I Error Rates as a Function of Method and Sample Size Ratio (α=0.05)

	n_e/n_c		
Method	.67	1.0	1.5
TT	.052	.049	.048
WT	.046	.049	.054
BT	.062	.066	.069
CT1	.053	.050	.046
CT2	.053	.049	.048
JT	.056	.052	.050

error rates. The Type I error rates for the CT1 and CT2 methods are slightly inflated when $n_e/n_c=0.67$ and slightly deflated when $n_e/n_c=1.5$. On the other hand, the Type I error rates for the WT method are slightly deflated when $n_e<n_c$ and slightly inflated when $n_e>n_c$. Type I error rates for the BT and WT methods increase as n_e/n_c increases, while for the other methods the opposite is the case. This is due to the effect of the lognormal distribution.

Power Analyses

In this section power results are presented in two categories: (a) power for variance differences only, and (b) power for mean and variance differences. In the first category, power is investigated when there is only a variance difference and the means of the two groups are equal. This is the case where $\mu_e=\mu_c$ and $\sigma_e\neq\sigma_c$. In the second category, power is investigated when there are mean and variance differences between the two groups. This is the case where $\mu_e\neq\mu_c$ and $\sigma_e\neq\sigma_c$.

Power for Variance Differences Only

In Table 13 a summary of the distribution of estimated power is presented, for conditions in which only the variances are different between the two groups. Of course the TT and WT methods are not designed to detect variance differences and some would view the power estimates for these tests as estimated Type I error rates. Results in the table show that power can be less than $\alpha=0.05$ for all

Table 13

Distribution of the Estimated Power (Variance Differences Only) for the Six Methods (α =0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.013	.026	.033	.054	.073	.087	.117
WT	.024	.030	.041	.050	.059	.074	.090
BT	.007	.015	.026	.101	.191	.236	.249
CT1	.000	.000	.003	.070	.391	.650	.899
CT2	.000	.001	.010	.045	.190	.461	.746
JT	.000	.001	.016	.044	.134	.333	.642

methods. Then there is little difference among the methods. When power is larger than α=0.05, the WT and TT methods have power near α, whereas the other methods have power above α. In general, the methods with power above α are ordered CT1, CT2, JT, and BT in terms of power. Tables 14 and 15 show the distributions of estimated power when data are from the normal and lognormal distributions, respectively. In general, sampling from a lognormal distribution appears to depress power for the CT1, CT2, JT, and BT methods.

Analysis of variance (ANOVA) was conducted to investigate the effect on the estimated power of the following factors: total sample size (N), sample size ratio (n_e/n_c), standard deviation ratio (σ_e/σ_c), distribution (D), and method (M). Because there were many significant main and interaction effects only the effects which contributed to more than 1% of the total variance were considered for interpretation purposes. Table 16 presents these effects and the percentage of variances. The results for power for tests conducted at α=0.01 and α=0.10 are similar to those reported in Table 16. Variance components for these tests are reported in Appendix B.

The σ_e/σ_c, M, and M$x\sigma_e/\sigma_c$ effects. Results for the M$x\sigma_e/\sigma_c$ interaction are shown in Table 17. The table shows that as expected, the alternative methods (BT, CT1, CT2, and JT) have low power when $\sigma_e<\sigma_c$. The alternative methods are designed for use when $\sigma_e \geq \sigma_c$ and consequently are not

Table 14

Distributions of Estimated Power (Variance Differences Only) for the Six Methods: Normal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.025	.029	.033	.053	.073	.083	.087
WT	.043	.046	.048	.050	.052	.054	.056
BT	.007	.014	.017	.079	.183	.234	.249
CT1	.000	.000	.000	.138	.574	.749	.899
CT2	.000	.000	.001	.079	.416	.577	.746
JT	.000	.000	.002	.067	.331	.482	.642

Table 15

Distributions of Estimated Power (Variance Differences Only) for the Six Methods: Lognormal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.013	.016	.032	.056	.072	.106	.117
WT	.024	.027	.031	.050	.072	.081	.090
BT	.021	.030	.041	.111	.199	.236	.243
CT1	.001	.002	.006	.070	.235	.367	.468
CT2	.006	.011	.020	.045	.103	.148	.240
JT	.012	.020	.031	.044	.065	.093	.145

Table 16

Percent of Total Variance for Estimated Power (Variance Differences Only) Effects Accounting ≥1% of the Variance (α=0.05)

Effect	Percentage
N	1.59
MxN	2.07
MxNxσ_e/σ_c	2.35
Nxσ_e/σ_c	2.42
D	4.34
Mxσ_e/σ_cxD	5.45
MxD	5.58
σ_e/σ_cxD	7.88
M	11.73
Mxσ_e/σ_c	24.54
σ_e/σ_c	28.01

Table 17

Mean Estimated Power (Variance Differences Only) as a Function of Method and Standard Deviation Ratio (α=0.05)

	σ_e/σ_c			
Method	1/3	1/2	2	3
TT	.068	.062	.043	.046
WT	.064	.060	.041	.040
BT	.027	.033	.154	.230
CT1	.004	.007	.337	.515
CT2	.009	.015	.215	.318
JT	.015	.022	.167	.242

expected to have good power when $\sigma_e<\sigma_c$ regardless of the differences between σ_e and σ_c. The alternative methods have better power when $\sigma_e>\sigma_c$, since the alternative methods are designed to detect $\sigma_e>\sigma_c$. Estimates for the traditional tests (TT and WT) are in actuality Type I error rates and reflect the effects of σ_e/σ_c on these rates reported in Algina, Oshima, and Lin (1994). Marginal estimated power means for the σ_e/σ_c effect are reported in Table 18. As expected from Table 17, results in the table indicate that power when $\sigma_e<\sigma_c$ is very low, and power improves as σ_e/σ_c increases. In Table 19, means of the estimated power for M are presented. Power estimates for the alternative methods are low because power is averaged for $\sigma_e<\sigma_c$ and $\sigma_e>\sigma_c$. Of the four alternative methods, the CT1 method appears to have the best power.

The MxD and Mxσ_e/σ_cxD effects. Table 20 gives results for the Mxσ_e/σ_cxD interaction. The behavior of TT and WT reflect the influence of σ_e/σ_c and D reported in Algina, Oshima, and Lin (1994). Power for the alternative methods is slightly lower for the lognormal distribution than for the normal distribution. An exception is the BT method which has higher power for the lognormal distribution than for the normal distribution. This trend follows the effect of lognormality on the Type I error rate for BT. Mean estimated power values for the MxD interaction are presented in Table 21. Following the results in Table 20, these results show

Table 18

Mean Estimated Power (Variance Differences Only) as a Function of Standard Deviation Ratio (α=0.05)

σ_e/σ_c	Means
1/3	.031
1/2	.033
2	.160
3	.232

Table 19

Mean Estimated Power (Variance Differences Only) as a Function of Method (α=0.05)

Method	Means
TT	.054
WT	.051
BT	.111
CT1	.216
CT2	.140
JT	.112

Table 20

Mean Estimated Power (Variance Differences Only) as a Function of Method, Standard Deviation Ratio and Distribution (α=0.05)

		σ_e/σ_c			
Disrtrb.	Method	1/3	1/2	2	3
N	TT	.054	.053	.051	.055
	WT	.050	.051	.049	.051
	BT	.015	.020	.143	.227
	CT1	.001	.002	.478	.704
	CT2	.001	.004	.341	.504
	JT	.002	.006	.277	.405
L	TT	.081	.071	.035	.036
	WT	.078	.069	.033	.030
	BT	.040	.045	.165	.233
	CT1	.007	.011	.196	.325
	CT2	.018	.026	.090	.133
	JT	.028	.039	.058	.079

Table 21

Mean Estimated Power (Variance Differences Only) as a Function of Method and Distribution (α=0.05)

	Distribution	
Method	N	L
TT	.053	.056
WT	.050	.052
BT	.101	.120
CT1	.296	.135
CT2	.212	.066
JT	.172	.051

that the CT1, CT2 and JT methods have better power when data are normally distributed than when the data are lognormally distributed. The BT method shows slightly higher power when data are from the lognormal distribution than when the data are from the normal distribution.

The MxN and MxNxσ_e/σ_c effects. In Table 22 estimated means for the MxNxσ_e/σ_c interaction are presented. These results indicate that when $\sigma_e<\sigma_c$, the alternative methods show low power, and the power becomes lower as the total sample size increases. When $\sigma_e>\sigma_c$, power for the alternative methods is higher, and improves as the total sample size increases. When $\sigma_e>\sigma_c$, the BT method appears not to be affected by the total sample size. But when $\sigma_e<\sigma_c$, the power for the BT method declines as the total sample size increases. Thus, the alternative methods are recommended when $\sigma_e>\sigma_c$. Table 23 presents results for the MxN interaction. Power for the alternative methods appear to be low for small total sample size, but improves as the total sample size increases. The only exception is the BT method which declines in power as the total sample size increases. As expected, the estimates reported for the TT and WT methods converges to α, as the total sample size increases.

The N and Nxσ_e/σ_c effects. Tables 24 and 25 display the results for interpreting the Nxσ_e/σ_c interaction and the N main effect. Results in Table 24 indicate that when $\sigma_e<\sigma_c$, power is low and becomes worse as the total sample size is

Table 22

Mean Estimated Power (Variance Differences Only) as a Function of Method, Sample Size and Standard Deviation Ratio (α=0.05)

		σ_e/σ_c			
Size	Method	1/3	1/2	2	3
40	TT	.072	.066	.042	.045
	WT	.068	.064	.041	.039
	BT	.031	.038	.154	.232
	CT1	.008	.013	.212	.341
	CT2	.016	.024	.124	.186
	JT	.024	.034	.102	.147
80	TT	.067	.060	.043	.045
	WT	.064	.058	.041	.039
	BT	.027	.030	.154	.230
	CT1	.002	.004	.347	.538
	CT2	.008	.012	.220	.328
	JT	.013	.019	.168	.245
120	TT	.064	.061	.043	.047
	WT	.061	.059	.041	.042
	BT	.024	.030	.154	.228
	CT1	.001	.003	.454	.666
	CT2	.004	.008	.302	.442
	JT	.009	.014	.232	.335

Table 23

Mean Estimated Power (Variance Differences Only) as a Function of Method and Total Sample Size (α=0.05)

	Size		
Method	40	80	120
TT	.056	.054	.054
WT	.053	.051	.051
BT	.114	.110	.109
CT1	.143	.223	.281
CT2	.087	.142	.189
JT	.077	.111	.147

Table 24

Mean Estimated Power (Variance Differences Only) as a Function of Sample Size and Standard Deviation Ratio (α=0.05)

	σ_e/σ_c			
Size	1/3	1/2	2	3
40	.036	.040	.112	.165
80	.030	.030	.162	.238
120	.027	.029	.204	.293

Table 25

Mean Estimated Power (Variance Differences Only) as a Function of Total Sample Size (α=0.05)

Size	Means
40	.088
80	.115
120	.138

increased. However, when $\sigma_e > \sigma_c$, power is moderate and improves as the total sample size is increased. Results for the total sample size are displayed in Table 25. The results do not provide any new information about the effect of N.

The D and D x σ_e/σ_c effects. Shown in Table 26 are means for interpreting the D x σ_e/σ_c interaction. As may be seen from the table, the results indicate that when $\sigma_e < \sigma_c$, power is generally low, but power for the lognormal case is better than for the normal case. However, when $\sigma_e > \sigma_c$, power is better, and power for the normal case is better than power for the lognormal case. In Table 27, the results of the distribution effect are shown. The results indicate that power is better when sampling is from the normal distribution than when sampling is from the lognormal distribution.

Power for Mean and Variance Differences

Table 28 presents a summary of the distribution of estimated power for the six methods, when there are mean and variance differences between the two groups. These results show that the alternative methods have better power than the traditional (TT and WT) methods. Tables 29 and 30 present a summary of the distributions of estimated power when data are normal and lognormal, respectively. These two tables show that the alternative methods have higher power when data are normal than

Table 26

Mean Estimated Power (Variance Differences Only) as a Function of Standard Deviation Ratio and Distribution (α=0.05)

	σ_e/σ_c			
Distrb	1/3	1/2	2	3
N	.020	.023	.223	.324
L	.081	.043	.096	.139

Table 27

Mean Estimated Power (Variance Differences Only) as a Function of Distribution (α=0.05)

Distrb.	Means
N	.148
L	.080

Table 28

Distribution of Estimated Power (Mean and Variance Differences)
(α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.091	.185	.242	.321	.428	.583	.644
WT	.111	.182	.244	.320	.420	.582	.660
BT	.051	.130	.204	.475	.601	.728	.814
CT1	.000	.008	.038	.336	.762	.939	.994
CT2	.000	.014	.074	.299	.636	.848	.961
JT	.000	.022	.104	.284	.590	.795	.936

Table 29

Distributions of Estimated Power (Mean and Variance Differences) for the Six Methods: Normal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.125	.185	.239	.318	.425	.579	.625
WT	.158	.181	.242	.309	.412	.574	.637
BT	.051	.093	.164	.453	.607	.732	.814
CT1	.000	.001	.014	.394	.909	.970	.994
CT2	.000	.004	.023	.357	.821	.914	.961
JT	.000	.006	.035	.349	.770	.871	.936

Table 30

Distributions of Estimated Power (Mean and Variance Differences) for the Six Methods: Lognormal Distribution (α=0.05)

		Percentile					
Method	MIN	10	25	50	75	90	MAX
TT	.091	.182	.253	.340	.432	.588	.644
WT	.111	.182	.249	.326	.431	.585	.660
BT	.089	.160	.250	.475	.601	.728	.791
CT1	.011	.036	.073	.337	.636	.759	.846
CT2	.028	.075	.136	.299	.460	.621	.747
JT	.048	.109	.179	.270	.403	.496	.684

when data are lognormal, but the opposite is the case for the traditional methods.

Analysis of variance (ANOVA) was conducted to investigate the effect on the power of the following factors: total sample size (N), sample size ratio (n_e/n_c), standard deviation ratio (σ_e/σ_c), distribution (D), mean differences/actual power (P), and method (M). Only the effects which contributed to more than 1% of the total variance were considered for interpretation. Table 31 shows the effects that accounted for more than 1% of the total variance. The results for power for tests conducted at α=0.01 and α=0.10 are similar to those reported in Table 31. Variance components for these tests are reported in Appendix B.

The σ_e/σ_c and M x σ_e/σ_c effects. Table 32 presents means for interpreting the M x σ_e/σ_c interaction. These results show that, except for the CT1 method, the methods achieve the highest power when $\sigma_e=\sigma_c$. The alternative methods are designed for use when $\sigma_e \geq \sigma_c$. As a result, when $\sigma_e \leq \sigma_c$, power for the traditional methods is better than power for the alternative methods. However, when $\sigma_e > \sigma_c$, the opposite is the case. Of the four alternative methods, when $\sigma_e > \sigma_c$, the CT1 method appears to have the highest power. Also, unlike the other methods, power for the CT1 method increases as the standard deviation ratio increases. The decline in power for the BT method is explained theoretically in Chapter 4. The

Table 31

Percent of Total Variance for Estimated Power (Mean and Variance Differences) Effects Accounting ≥1% of the Variance (α=0.05)

Effect	Percentage
σ_e/σ_cxP	2.10
Mxσ_e/σ_cxD	2.56
σ_e/σ_cxD	4.08
P	12.16
Mxσ_e/σ_c	22.36
σ_e/σ_c	50.84

Table 32

Mean Estimated Power (Mean and Variance Differences) as a Function of Method and Standard Deviation Ratio (α=0.05)

	σ_e/σ_c				
Method	1/3	1/2	1	2	3
TT	.293	.433	.704	.413	.252
WT	.292	.433	.703	.412	.249
BT	.157	.294	.689	.634	.596
CT1	.028	.088	.569	.739	.773
CT2	.056	.146	.634	.651	.609
JT	.078	.190	.670	.603	.526

TT and WT methods show similar power regardless of the standard deviation ratio. Shown in Table 33 are the results of the marginal mean estimated power for the σ_e/σ_c effect. The results indicate that power is low for the negative condition ($\sigma_e<\sigma_c$) and moderate for the positive condition ($\sigma_e\geq\sigma_c$). For the positive condition, power declines as the standard deviation ratio increases.

The σ_e/σ_cxD and M$x\sigma_e/\sigma_c$xD effects. Table 34 shows the cell means for the M$x\sigma_e/\sigma_c$xD interaction. When $\sigma_e<\sigma_c$, as expected, the traditional methods have higher power than the alternative methods, for both normal and lognormal data. This is because alternative methods were not designed for the case $\sigma_e<\sigma_c$. When $\sigma_e=\sigma_c$, the traditional are still more powerful than the alternative methods, with the TT method showing the highest power. It is well known that from theory that, when $\sigma_e=\sigma_c$ and data is normal, the TT method is a uniformly powerful method. When $\sigma_e>\sigma_c$, as expected, the alternative methods are more powerful than the traditional methods, with the CT1 method showing the highest power. This is because the alternative methods are designed for use when $\sigma_e>\sigma_c$. In terms of power, when $\sigma_e>\sigma_c$, the order of the alternative methods is CT1, CT2, BT and JT. Results in the table also show that when $\sigma_e\leq\sigma_c$, the TT,WT, BT and JT methods have better power when data are generated from the lognormal distribution than when data are generated from the normal distribution. However, when $\sigma_e>\sigma_c$, all the methods

Table 33

Mean Estimated Power (Mean and Variance Differences) as a Function of Standard Deviation Ratio (α=0.05)

σ_e/σ_c	Means
1/3	.151
1/2	.264
1	.662
2	.576
3	.501

Table 34

Mean Estimated Power (Mean and Variance Differences) as a Function of Method, Standard Deviation Ratio and Distribution (α=0.05)

		σ_e/σ_c				
D	M	1/3	1/2	1	2	3
N	TT	.268	.416	.700	.416	.266
	WT	.265	.415	.698	.415	.265
	BT	.122	.262	.692	.637	.601
	CT1	.009	.047	.605	.859	.901
	CT2	.015	.071	.640	.792	.792
	JT	.022	.097	.663	.752	.725
L	TT	.319	.450	.709	.411	.237
	WT	.318	.451	.708	.410	.233
	BT	.193	.325	.686	.631	.592
	CT1	.048	.128	.533	.618	.646
	CT2	.096	.220	.628	.511	.427
	JT	.134	.284	.678	.454	.328

have better power when data are normal than when data are lognormal. For all the methods, when $\sigma_e=\sigma_c$, the power difference between distributions is very small. In Table 35 results of the σ_e/σ_cxD interaction are displayed. These results indicate that when $\sigma_e<\sigma_c$, power is better when data are lognormal than when data are normal, but the opposite is the case when $\sigma_e\geq\sigma_c$.

The P and P$x\sigma_e/\sigma_c$ effects. Table 36 shows the means for interpreting the P$x\sigma_e/\sigma_c$ interaction. The results indicate that qualitatively the influence of σ_e/σ_c on power is the same for all levels of target power. Also, estimated power increases as the target power increases. Table 37 shows the means for interpreting the P main effect. The table shows that estimated power increases as the target power increases, but the estimated power is much lower than the target power. This is due to the fact that the estimated power is averaged over other factors, some of which have a negative impact on the estimated power.

Summary

Results of this study have shown that:

1. By Bradley's (1978) liberal criterion, all the methods investigated maintained good control of the Type I error rates,

except for the BT method which exhibited substantially larger Type I error rates than the other methods when data were lognormal.

Table 35

Mean Estimated Power (Mean and Variance Differences) as a Function of Standard Deviation Ratio and Distribution(α=0.05)

	σ_e/σ_c				
D	1/3	1/2	1	2	3
N	.117	.218	.666	.645	.592
L	.185	.310	.657	.506	.411

Table 36

Mean Estimated Power (Mean and Variance Differences) as a Function of Standard Deviation Ratio and Actual Power(α=0.05)

	σ_e/σ_c				
P	1/3	1/2	1	2	3
.5	.104	.171	.471	.445	.417
.7	.141	.248	.658	.559	.489
.9	.207	.373	.858	.723	.593

Table 37

Mean Estimated Power (Mean and Variance Difference) as a Function of Actual Power (α=0.05)

Power	Means
.5	.322
.7	.419
.9	.551

2. In general, all the methods performed better, in terms of Type I error rates and power, when the data were generated from the normal distribution than when the data were generated from the lognormal distribution.

3. When $\sigma_e<\sigma_c$, the power was low and the traditional methods were more powerful than the alternative methods. When $\sigma_e=\sigma_c$, the the power was higher and the traditional methods were still more powerful than the alternative methods. Finally, when $\sigma_e>\sigma_c$, the alternative methods were more powerful than the traditional methods.

CHAPTER 4
CONCLUSIONS

The purpose of this study was to investigate Type I error rates and power of some alternatives to the independent samples t-test when a difference in means is accompanied by a difference in variances. Three hundred and sixty simulated conditions were used to investigate the effects of several factors on the Type I error rates and power of the methods. The results of the study are intended to provide researchers with some guidelines for comparing the means of two populations when a difference in means is accompanied by a difference in variances. However, the generalizability of the results is limited by the range of values of the total sample size, sample size ratio, standard deviation ratio, mean difference and distribution.

Conclusions

Within the limitations of the study, the following findings of the present study can be made:

Conclusion 1. All the methods maintained control of the Type I error rates, except for the BT method, which does not control the Type I error rate under lognormality.

Conclusion 2. All the alternative methods had smaller

Type I error rates when $n_e > n_c$ than when $n_e < n_c$, except BT in which the opposite was true.

Conclusion 3. When $\sigma_e > \sigma_c$, the alternative methods had better power than the traditional methods, but when $\sigma_e < \sigma_c$, the opposite was true. This was true for both the case when only a variance difference existed between the two groups, and when there was a mean and variance difference between the two groups. Among the alternative methods, CT1 had the best power.

Conclusion 4. For all the methods, except for CT1, power was the highest when $\sigma_e = \sigma_c$ and $\mu_e \neq \mu_c$. In addition, when $\sigma_e = \sigma_c$ and $\mu_e \neq \mu_c$, power for the TT and WT methods was better than power for the alternative methods.

Conclusion 5. Power was better when the data were sampled from the normal distribution than when the data was sampled from the lognormal distribution.

Conclusion 6. When $\sigma_e \geq \sigma_c$, power for all the alternative methods except for the CT1 method, declined as the standard deviation ratio increased. This suggests that the alternative methods perform well to a certain degree of the differences between the variances of the two groups. However, it should be noted that when $\sigma_e > \sigma_c$ the power of the traditional methods declined more sharply than did the alternative methods.

Theoretical Explanations of Selected Results

Explanation of Conclusion 1 for BT Method

Results in this study have shown that the BT method has inflated Type I error rates when data are sampled from the lognormal distribution. According to Algina, Oshima, and Lin (1994), for non-normal data, the Type I error rate depends on the sampling covariance between the numerator and the square of the denominator for the test statistic. When the covariance is not equal to zero, extreme mean differences are associated with small standard errors. For the one tailed test, as in this study, the Type I error rate depends on the sign of the sampling covariance and the tail of the test. Upper tailed tests tend to have estimated Type I error rate larger than the nominal rate when the sampling covariance is negative, and estimated Type I error rate smaller than the nominal rate when the sampling covariance is positive.

From Kendall and Stuart (1973), the sampling covariance between the mean and variance for a single sample is μ_3/n, where μ_3 is the third central moment and n is the total sample size. Using this result, from Equation 35, the covariance between the numerator and the square of the denominator for the BT method is given by

$$\sigma_{BT^2} = -\frac{\mu_3}{n_c}\left(\frac{1}{n_e}+\frac{1}{n_c}\right) = -\frac{\gamma_3 \sigma_c^3}{n_c}\left(\frac{1}{n_e}+\frac{1}{n_c}\right) \qquad (39)$$

where $\gamma_3=\mu_3/\sigma^3$. Since σ_{BT} is negative and the test is upper tailed, the estimated Type I error rate is larger than the nominal significance level. This explains why the BT method has inflated Type I error rates when data is sampled from the lognormal distribution, as shown in Table 11.

Explanation of Conclusion 2 for BT Method

Algina, Oshima, and Lin (1994) showed that the discrepancies between the estimated Type I error rate and the nominal significance level are larger for larger sampling covariance. For fixed N, $1/n_c(1/n_c + 1/n_e)$ is smaller when $n_e>n_c$ than when $n_e=n_c$ or $n_c>n_e$. Therefore, as n_e/n_c increases, σ_{BT} also increases and hence the estimated Type I error rate also increases. This explains the increase in Type I error rate for the BT method as n_e/n_c increases, as shown in Table 12.

Explanation of Conclusion 6 for BT Method

For conditions in which $\sigma_e<\sigma_c$, $\sigma_e=1.0$ and σ_c increases as σ_e/σ_c declines. From Equation 35, the denominator for BT is

$$\sqrt{\sigma_c(\frac{1}{n_c}+\frac{1}{n_e})}\,.$$

Thus as σ_e/σ_c declines the denominator of BT increases and consequently power declines. For conditions in which $\sigma_e>\sigma_c$, σ_c =1.0. In Equation 35, notice that the denominator is not a function of the standard deviation for the experimental group, σ_e. However, Brownie, Boos and Hughes-Oliver (1990) have shown that, under normality, the power function for BT is

$$Pr[T(n_c-1,\delta)>[\frac{n_c+n_e}{n_c+\rho n_e}]^{1/2}t(\alpha;n_c-1)],$$

where $\rho=\sigma_e^2/\sigma_c^2$, and $t(\alpha;n_c-1)$ is the 1-α percentile of Student's t distribution with (n_c-1) degrees of freedom and

$$\delta=\frac{\frac{\mu_e-\mu_c}{\sigma_c}}{[\frac{1}{n_c}+\frac{\sigma_e^2}{\sigma_c^2}\frac{1}{n_e}]^{1/2}}.$$

Calculations show that for conditions in which power for TT was .50, under normality, $\sigma_e/\sigma_c\geq1.0$ has little effect on the power for BT. For conditions in which power for TT was .70 or .90, under normality, as $\sigma_e/\sigma_c\geq1.0$ increased, power for BT declined.

Comparisons to Previous Studies

Conover and Salsburg (1988) used several sets of data from biometrics experiments to compare several methods including the TT, CT1 and CT2 methods. In each case, their

results showed that the CT1 and CT2 methods had better power than the TT method. Results in this study are consistent with the findings of Conover and Salsburg. As can be seen in Table 32, when $\sigma_e > \sigma_c$, the two methods suggested by Conover and Salsburg are superior to the independent samples t-test.

Brownie et al. (1990) compared the TT method and the BT method. Using power plots, they showed that

(1) Power for the BT method was better than the TT method.

(2) Power gains increased as the standard deviation ratio increased.

(3) Power gains increased at a faster rate for $n_e = n_c$ than for $n_e \neq n_c$.

Results of this study also verified the superiority of the BT method over the TT method (see Table 32). As can be seen in Table 32, power gains increased with an increase in the standard deviation ratio. Thus, the results from this study are consistent with the results from the Brownie et al. study.

As noted in Chapter 1, Blair and Sawilowsky (1993) found that, for non-normal data, the TT method was more powerful than the BT method, when $\sigma_e = \sigma_c$. However, when $\sigma_e > \sigma_c$, the BT method was more powerful than the TT method. Also, for non-normal data, the TT method had better control of the Type I error rate than the BT method. Results in the study are consistent with the results from Blair and Sawilowsky.

Results in Table 6 show that, when data are lognormal, the TT method has better control of the Type I error rate than the BT method. In terms of power, Table 34 shows that, when $\sigma_e=\sigma_c$ and data are lognormal, the TT method is more powerful than the BT method. The table also shows that, when $\sigma_e>\sigma_c$, the BT method is more powerful than the TT method.

Johnson at al. (1987) showed that the JT method had control of Type I error rates. Also, in their study, they showed that power increased as the total sample size increased. Table 23 shows that the power increases with an increase in total sample size. Thus, results of this study are consistent with the Johnson et al. findings.

Suggestions for Future Research

As mentioned above, generalizability of the results in this study limited by the range of values for (a) total sample size, (b) distribution, (c) standard deviation ratio, (d) mean difference and (e) sample size ratio. Results in the study have shown that the total sample size did not have a significant impact on power of the methods investigated in the study, when there is a mean and variance difference between the two groups. Also, in the literature, most of the alternative methods were applied to relatively small total sample size. Thus, further research needs to be conducted to see how the Type I error rates and power of the methods are affected when the total sample size is small.

With one exception, results in the study have shown that as the standard deviation increases, power for the alternative methods starts to decline. As Stevens (1979) and Wilcox (1987) have shown, data in which the standard deviation ratios are larger than those investigated in this study, do occur in practice. Thus, further research is needed to examine the extent to which the standard deviation ratio can be increased before the alternative methods fail.

This study included only data sampled from the normal and lognormal distributions. In general, results from the study showed that the methods did not perform well for the lognormal distribution. Even though Miceri's (1989) results gave evidence that the distributions in this study are not unrepresentative of distributions found in practice, further research in this area is needed to examine the effect of other non-normal distributions on the power of the alternative methods. In particular, further research is needed to investigate the effects of skewness and kurtosis on the alternative methods.

Finally, in this study, for the analysis of the JT method, Δ/σ was set to 0.6. However, as Table 38 shows, the power of JT, as a function of σ_e/σ_c, is affected by the choice of Δ/σ. When $\sigma_e=\sigma_c$, the highest power is achieved when $\Delta/\sigma=.2$. However, for $\sigma_e>\sigma_c$, high power is achieved at higher levels of Δ/σ. This suggests that, in order to achieve better power when using the JT, researchers might

Table 38

Mean Estimated Power (Mean and Variance Differences) for the JT method as Function of Target Power, Standard Deviation Ratio and Δ/σ (α=0.05)

		σ_e/σ_c				
P	Δ/σ	1/3	1/2	1	2	3
.5	.2	.095	.179	.502	.353	.247
	.4	.050	.123	.489	.444	.364
	.6	.027	.078	.480	.522	.476
	.8	.016	.055	.460	.581	.576
	1	.009	.039	.440	.638	.662
.7	.2	.130	.271	.682	.448	.333
	.4	.070	.201	.672	.541	.456
	.6	.038	.137	.652	.623	.560
	.8	.021	.095	.631	.692	.652
	1	.013	.072	.600	.739	.734
.9	.2	.212	.440	.894	.640	.442
	.4	.133	.343	.883	.721	.561
	.6	.080	.255	.867	.783	.670
	.8	.047	.187	.847	.833	.760
	1	.030	.137	.821	.865	.825

have to use different values of Δ/σ depending on the standard deviation ration. Thus, further research is needed to investigate combinations of Δ/σ and σ_e/σ_c for the best power.

APPENDIX A
ESTIMATED TYPE I ERROR RATES AND POWER

Table 39

Estimated Type I error rates and Power When α=0.05

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	1.00	0.33	0.0	N	.052	.052	.015	.002	.004	.005
40	1.00	0.50	0.0	N	.054	.052	.023	.008	.012	.015
40	1.00	1.00	0.0	N	.049	.049	.052	.051	.049	.054
40	1.00	2.00	0.0	N	.055	.054	.143	.299	.193	.163
40	1.00	3.00	0.0	N	.056	.053	.234	.493	.301	.246
40	1.00	0.33	0.5	N	.184	.179	.069	.011	.020	.028
40	1.00	0.50	0.5	N	.276	.273	.149	.043	.065	.084
40	1.00	1.00	0.5	N	.508	.507	.497	.415	.451	.486
40	1.00	2.00	0.5	N	.280	.276	.493	.659	.544	.507
40	1.00	3.00	0.5	N	.188	.181	.486	.730	.545	.471
40	1.00	0.33	0.7	N	.237	.229	.106	.018	.030	.043
40	1.00	0.50	0.7	N	.403	.395	.240	.073	.109	.142
40	1.00	1.00	0.7	N	.692	.692	.676	.582	.626	.656
40	1.00	2.00	0.7	N	.383	.377	.613	.746	.658	.624
40	1.00	3.00	0.7	N	.240	.232	.585	.793	.622	.560
40	1.00	0.33	0.9	N	.384	.372	.191	.036	.058	.078
40	1.00	0.50	0.9	N	.575	.570	.393	.150	.213	.258
40	1.00	1.00	0.9	N	.899	.899	.892	.804	.805	.869
40	1.00	2.00	0.9	N	.579	.574	.783	.867	.806	.787
40	1.00	3.00	0.9	N	.372	.361	.709	.865	.725	.671
40	0.67	0.33	0.0	N	.029	.053	.007	.001	.001	.002
40	0.67	0.50	0.0	N	.033	.052	.014	.003	.004	.006

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	0.67	1.00	0.0	N	.051	.051	.047	.054	.056	.061
40	0.67	2.00	0.0	N	.070	.045	.145	.310	.217	.193
40	0.67	3.00	0.0	N	.083	.046	.236	.494	.339	.284
40	0.67	0.33	0.5	N	.130	.191	.056	.009	.016	.022
40	0.67	0.50	0.5	N	.233	.301	.121	.044	.060	.073
40	0.67	1.00	0.5	N	.492	.490	.469	.420	.454	.481
40	0.67	2.00	0.5	N	.321	.251	.483	.654	.570	.537
40	0.67	3.00	0.5	N	.241	.161	.492	.725	.587	.523
40	0.67	0.33	0.7	N	.206	.291	.095	.018	.028	.040
40	0.67	0.50	0.7	N	.342	.425	.216	.075	.107	.133
40	0.67	1.00	0.7	N	.699	.694	.670	.603	.642	.672
40	0.67	2.00	0.7	N	.436	.348	.601	.738	.675	.646
40	0.67	3.00	0.7	N	.319	.231	.578	.788	.669	.610
40	0.67	0.33	0.9	N	.301	.403	.153	.034	.051	.065
40	0.67	0.50	0.9	N	.540	.628	.363	.154	.207	.246
40	0.67	1.00	0.9	N	.905	.904	.892	.830	.860	.878
40	0.67	2.00	0.9	N	.613	.520	.764	.862	.817	.796
40	0.67	3.00	0.9	N	.440	.332	.694	.849	.751	.701
40	1.50	0.33	0.0	N	.087	.047	.022	.003	.004	.007
40	1.50	0.50	0.0	N	.074	.049	.026	.007	.011	.016

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	1.50	1.00	0.0	N	.053	.052	.054	.047	.047	.054
40	1.50	2.00	0.0	N	.032	.049	.132	.267	.145	.118
40	1.50	3.00	0.0	N	.029	.055	.218	.452	.221	.171
40	1.50	0.33	0.5	N	.249	.166	.092	.015	.027	.038
40	1.50	0.50	0.5	N	.315	.243	.165	.044	.076	.098
40	1.50	1.00	0.5	N	.507	.503	.501	.405	.446	.480
40	1.50	2.00	0.5	N	.228	.299	.500	.631	.497	.454
40	1.50	3.00	0.5	N	.138	.203	.503	.716	.481	.401
40	1.50	0.33	0.7	N	.311	.224	.132	.018	.041	.057
40	1.50	0.50	0.7	N	.438	.349	.250	.083	.127	.297
40	1.50	1.00	0.7	N	.693	.688	.685	.567	.619	.652
40	1.50	2.00	0.7	N	.332	.410	.631	.742	.620	.579
40	1.50	3.00	0.7	N	.189	.256	.594	.780	.560	.487
40	1.50	0.33	0.9	N	.421	.317	.209	.040	.068	.092
40	1.50	0.50	0.9	N	.625	.536	.423	.156	.232	.284
40	1.50	1.00	0.9	N	.893	.888	.881	.769	.824	.854
40	1.50	2.00	0.9	N	.543	.629	.804	.868	.789	.759
40	1.50	3.00	0.9	N	.302	.411	.732	.863	.695	.634
80	1.00	0.33	0.0	N	.053	.052	.014	.000	.000	.001
80	1.00	0.50	0.0	N	.047	.047	.020	.002	.003	.003

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
80	1.00	1.00	0.0	N	.053	.053	.051	.050	.051	.053
80	1.00	2.00	0.0	N	.043	.043	.142	.494	.353	.276
80	1.00	3.00	0.0	N	.048	.047	.228	.749	.527	.417
80	1.00	0.33	0.5	N	.176	.173	.066	.001	.002	.004
80	1.00	0.50	0.5	N	.276	.274	.157	.012	.023	.034
80	1.00	1.00	0.5	N	.507	.507	.499	.423	.454	.475
80	1.00	2.00	0.5	N	.271	.269	.487	.824	.731	.665
80	1.00	3.00	0.5	N	.179	.176	.496	.910	.779	.687
80	1.00	0.33	0.7	N	.249	.245	.108	.004	.006	.010
80	1.00	0.50	0.7	N	.398	.395	.247	.026	.046	.062
80	1.00	1.00	0.7	N	.693	.693	.685	.585	.629	.650
80	1.00	2.00	0.7	N	.402	.399	.631	.888	.824	.771
80	1.00	3.00	0.7	N	.249	.246	.588	.943	.836	.756
80	1.00	0.33	0.9	N	.370	.366	.183	.006	.009	.016
80	1.00	0.50	0.9	N	.595	.593	.414	.067	.103	.131
80	1.00	1.00	0.9	N	.905	.905	.900	.817	.854	.870
80	1.00	2.00	0.9	N	.586	.583	.796	.953	.916	.882
80	1.00	3.00	0.9	N	.379	.375	.719	.968	.902	.844
80	0.67	0.33	0.0	N	.026	.050	.009	.000	.000	.000
80	0.67	0.50	0.0	N	.034	.052	.014	.001	.002	.002
80	0.67	1.00	0.0	N	.048	.048	.047	.051	.048	.052

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
80	0.67	2.00	0.0	N	.073	.050	.152	.505	.397	.328
80	0.67	3.00	0.0	N	.083	.050	.244	.742	.577	.482
80	0.67	0.33	0.5	N	.145	.211	.060	.001	.003	.004
80	0.67	0.50	0.5	N	.223	.284	.130	.013	.019	.027
80	0.67	1.00	0.5	N	.492	.488	.487	.425	.455	.476
80	0.67	2.00	0.5	N	.324	.252	.496	.811	.744	.687
80	0.67	3.00	0.5	N	.233	.158	.497	.889	.791	.714
80	0.67	0.33	0.7	N	.186	.267	.080	.001	.004	.006
80	0.67	0.50	0.7	N	.349	.430	.220	.024	.037	.050
80	0.67	1.00	0.7	N	.697	.695	.686	.607	.640	.662
80	0.67	2.00	0.7	N	.440	.360	.620	.876	.827	.784
80	0.67	3.00	0.7	N	.292	.211	.570	.922	.843	.773
80	0.67	0.33	0.9	N	.314	.419	.156	.006	.009	.014
80	0.67	0.50	0.9	N	.533	.616	.365	.059	.079	.106
80	0.67	1.00	0.9	N	.903	.900	.896	.829	.859	.875
80	0.67	2.00	0.9	N	.609	.534	.767	.939	.914	.900
80	0.67	3.00	0.9	N	.415	.317	.691	.950	.892	.843
80	1.50	0.33	0.0	N	.082	.053	.023	.000	.000	.001
80	1.50	0.50	0.0	N	.073	.049	.024	.001	.003	.004
80	1.50	1.00	0.0	N	.050	.051	.052	.046	.049	.050
80	1.50	2.00	0.0	N	.030	.050	.135	.476	.292	.223

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
80	1.50	3.00	0.0	N	.032	.051	.207	.745	.461	.333
80	1.50	0.33	0.5	N	.241	.166	.093	.004	.007	.009
80	1.50	0.50	0.5	N	.320	.251	.167	.016	.028	.042
80	1.50	1.00	0.5	N	.495	.494	.490	.391	.424	.448
80	1.50	2.00	0.5	N	.237	.303	.504	.820	.690	.616
80	1.50	3.00	0.5	N	.129	.194	.497	.911	.728	.617
80	1.50	0.33	0.7	N	.310	.226	.133	.003	.008	.013
80	1.50	0.50	0.7	N	.430	.350	.247	.027	.049	.069
80	1.50	1.00	0.7	N	.690	.689	.685	.570	.608	.635
80	1.50	2.00	0.7	N	.340	.420	.640	.890	.793	.729
80	1.50	3.00	0.7	N	.195	.281	.591	.942	.794	.696
80	1.50	0.33	0.9	N	.411	.314	.200	.008	.014	.026
80	1.50	0.50	0.9	N	.610	.538	.416	.064	.107	.143
80	1.50	1.00	0.9	N	.897	.896	.896	.794	.835	.854
80	1.50	2.00	0.9	N	.539	.622	.798	.949	.898	.865
80	1.50	3.00	0.9	N	.299	.413	.726	.969	.885	.809
120	1.00	0.33	0.0	N	.054	.054	.016	.000	.000	.000
120	1.00	0.50	0.0	N	.057	.056	.021	.000	.000	.001
120	1.00	1.00	0.0	N	.050	.050	.051	.051	.051	.053
120	1.00	2.00	0.0	N	.048	.047	.146	.660	.496	.405
120	1.00	3.00	0.0	N	.053	.052	.222	.899	.710	.576

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	1.00	0.33	0.5	N	.178	.175	.066	.000	.000	.001
120	1.00	0.50	0.5	N	.282	.281	.158	.004	.007	.012
120	1.00	1.00	0.5	N	.496	.495	.493	.411	.436	.459
120	1.00	2.00	0.5	N	.272	.271	.499	.910	.841	.782
120	1.00	3.00	0.5	N	.185	.182	.500	.971	.894	.816
120	1.00	0.33	0.7	N	.253	.252	.115	.000	.001	.002
120	1.00	0.50	0.7	N	.384	.383	.238	.010	.020	.029
120	1.00	1.00	0.7	N	.708	.708	.703	.604	.643	.660
120	1.00	2.00	0.7	N	.403	.401	.636	.950	.900	.900
120	1.00	3.00	0.7	N	.243	.240	.599	.981	.930	.871
120	1.00	0.33	0.9	N	.367	.363	.185	.001	.002	.005
120	1.00	0.50	0.9	N	.587	.585	.407	.027	.046	.064
120	1.00	1.00	0.9	N	.892	.892	.890	.813	.843	.856
120	1.00	2.00	0.9	N	.583	.582	.792	.976	.958	.935
120	1.00	3.00	0.9	N	.367	.365	.724	.990	.961	.928
120	0.67	0.33	0.0	N	.025	.044	.009	.000	.000	.000
120	0.67	0.50	0.0	N	.033	.050	.014	.000	.001	.001
120	0.67	1.00	0.0	N	.048	.048	.047	.054	.053	.055
120	0.67	2.00	0.0	N	.071	.048	.160	.650	.537	.450
120	0.67	3.00	0.0	N	.086	.053	.249	.872	.746	.642
120	0.67	0.33	0.5	N	.125	.192	.051	.000	.000	.000

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	0.67	0.50	0.5	N	.229	.292	.136	.004	.007	.010
120	0.67	1.00	0.5	N	.501	.500	.491	.427	.453	.467
120	0.67	2.00	0.5	N	.320	.255	.504	.900	.848	.795
120	0.67	3.00	0.5	N	.232	.163	.504	.960	.906	.844
120	0.67	0.33	0.7	N	.188	.272	.087	.001	.001	.002
120	0.67	0.50	0.7	N	.352	.432	.223	.008	.014	.021
120	0.67	1.00	0.7	N	.709	.704	.702	.622	.652	.669
120	0.67	2.00	0.7	N	.442	.362	.617	.934	.903	.863
120	0.67	3.00	0.7	N	.320	.231	.587	.970	.926	.874
120	0.67	0.33	0.9	N	.305	.414	.163	.000	.002	.004
120	0.67	0.50	0.9	N	.547	.631	.396	.030	.054	.062
120	0.67	1.00	0.9	N	.901	.901	.900	.822	.849	.849
120	0.67	2.00	0.9	N	.609	.533	.775	.973	.958	.936
120	0.67	3.00	0.9	N	.428	.332	.693	.982	.956	.925
120	1.50	0.33	0.0	N	.079	.048	.018	.000	.000	.000
120	1.50	0.50	0.0	N	.075	.050	.026	.000	.001	.002
120	1.50	1.00	0.0	N	.046	.046	.048	.045	.050	.049
120	1.50	2.00	0.0	N	.033	.052	.134	.642	.435	.333
120	1.50	3.00	0.0	N	.029	.049	.206	.894	.656	.496
120	1.50	0.33	0.5	N	.245	.167	.090	.000	.001	.002
120	1.50	0.50	0.5	N	.316	.249	.167	.004	.009	.016

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	1.50	1.00	0.5	N	.506	.503	.500	.400	.432	.453
120	1.50	2.00	0.5	N	.231	.304	.508	.907	.809	.733
120	1.50	3.00	0.5	N	.127	.195	.504	.973	.864	.768
120	1.50	0.33	0.7	N	.317	.231	.130	.001	.002	.003
120	1.50	0.50	0.7	N	.434	.363	.254	.014	.023	.036
120	1.50	1.00	0.7	N	.712	.711	.707	.593	.638	.655
120	1.50	2.00	0.7	N	.356	.436	.650	.953	.896	.838
120	1.50	3.00	0.7	N	.189	.282	.616	.986	.915	.835
120	1.50	0.33	0.9	N	.433	.337	.213	.003	.005	.011
120	1.50	0.50	0.9	N	.624	.539	.419	.037	.059	.084
120	1.50	1.00	0.9	N	.898	.900	.900	.800	.842	.859
120	1.50	2.00	0.9	N	.556	.637	.814	.982	.958	.931
120	1.50	3.00	0.9	N	.306	.416	.742	.994	.954	.909
40	1.00	0.33	0.0	L	.090	.088	.049	.014	.030	.043
40	1.00	0.50	0.0	L	.072	.071	.051	.020	.037	.051
40	1.00	1.00	0.0	L	.051	.049	.097	.051	.049	.057
40	1.00	2.00	0.0	L	.033	.032	.169	.130	.063	.046
40	1.00	3.00	0.0	L	.028	.025	.236	.205	.082	.057
40	1.00	0.33	0.5	L	.255	.250	.145	.052	.092	.129
40	1.00	0.50	0.5	L	.336	.332	.241	.115	.189	.233
40	1.00	1.00	0.5	L	.523	.522	.515	.356	.442	.494

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	1.00	2.00	0.5	L	.245	.238	.505	.411	.316	.284
40	1.00	3.00	0.5	L	.147	.138	.494	.430	.255	.202
40	1.00	0.33	0.7	L	.321	.315	.202	.082	.134	.182
40	1.00	0.50	0.7	L	.453	.449	.336	.170	.272	.331
40	1.00	1.00	0.7	L	.705	.704	.671	.508	.622	.676
40	1.00	2.00	0.7	L	.380	.372	.617	.514	.439	.411
40	1.00	3.00	0.7	L	.204	.194	.581	.504	.320	.265
40	1.00	0.33	0.9	L	.441	.432	.284	.121	.203	.254
40	1.00	0.50	0.9	L	.605	.597	.465	.258	.387	.452
40	1.00	1.00	0.9	L	.891	.890	.846	.707	.807	.851
40	1.00	2.00	0.9	L	.614	.606	.777	.688	.642	.634
40	1.00	3.00	0.9	L	.348	.336	.701	.633	.467	.407
40	0.67	0.33	0.0	L	.059	.080	.031	.010	.021	.031
40	0.67	0.50	0.0	L	.058	.068	.042	.018	.034	.046
40	0.67	1.00	0.0	L	.056	.040	.087	.053	.055	.060
40	0.67	2.00	0.0	L	.048	.026	.162	.142	.079	.063
40	0.67	3.00	0.0	L	.059	.024	.238	.238	.125	.093
40	0.67	0.33	0.5	L	.207	.261	.122	.052	.091	.118
40	0.67	0.50	0.5	L	.282	.332	.194	.111	.160	.201
40	0.67	1.00	0.5	L	.520	.516	.505	.372	.456	.508
40	0.67	2.00	0.5	L	.295	.200	.486	.415	.341	.314

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	0.67	3.00	0.5	L	.196	.111	.477	.444	.297	.244
40	0.67	0.33	0.7	L	.260	.325	.166	.073	.125	.156
40	0.67	0.50	0.7	L	.396	.456	.288	.168	.246	.294
40	0.67	1.00	0.7	L	.707	.720	.664	.546	.633	.683
40	0.67	2.00	0.7	L	.428	.327	.601	.532	.466	.444
40	0.67	3.00	0.7	L	.268	.167	.570	.521	.372	.317
40	0.67	0.33	0.9	L	.391	.466	.261	.122	.190	.236
40	0.67	0.50	0.9	L	.573	.638	.438	.288	.392	.455
40	0.67	1.00	0.9	L	.893	.908	.848	.759	.840	.875
40	0.67	2.00	0.9	L	.639	.556	.753	.691	.662	.656
40	0.67	3.00	0.9	L	.412	.275	.682	.632	.494	.444
40	1.50	0.33	0.0	L	.117	.089	.061	.017	.035	.054
40	1.50	0.50	0.0	L	.106	.090	.069	.024	.047	.068
40	1.50	1.00	0.0	L	.045	.060	.096	.047	.046	.051
40	1.50	2.00	0.0	L	.017	.038	.173	.116	.044	.032
40	1.50	3.00	0.0	L	.015	.031	.229	.162	.046	.031
40	1.50	0.33	0.5	L	.299	.234	.176	.054	.105	.143
40	1.50	0.50	0.5	L	.377	.325	.268	.108	.188	.249
40	1.50	1.00	0.5	L	.532	.534	.535	.341	.438	.505
40	1.50	2.00	0.5	L	.199	.293	.519	.385	.280	.250
40	1.50	3.00	0.5	L	.091	.159	.505	.393	.187	.142

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
40	1.50	0.33	0.7	L	.376	.306	.289	.080	.147	.191
40	1.50	0.50	0.7	L	.488	.422	.352	.160	.265	.331
40	1.50	1.00	0.7	L	.717	.706	.689	.481	.611	.675
40	1.50	2.00	0.7	L	.342	.439	.640	.487	.406	.381
40	1.50	3.00	0.7	L	.141	.230	.587	.466	.257	.208
40	1.50	0.33	0.9	L	.493	.411	.319	.110	.203	.260
40	1.50	0.50	0.9	L	.644	.575	.481	.231	.369	.446
40	1.50	1.00	0.9	L	.900	.880	.853	.672	.802	.851
40	1.50	2.00	0.9	L	.580	.660	.791	.666	.621	.613
40	1.50	3.00	0.9	L	.274	.408	.728	.608	.409	.343
80	1.00	0.33	0.0	L	.081	.080	.037	.001	.015	.024
80	1.00	0.50	0.0	L	.065	.065	.040	.001	.022	.034
80	1.00	1.00	0.0	L	.046	.046	.076	.049	.049	.050
80	1.00	2.00	0.0	L	.037	.037	.166	.207	.093	.059
80	1.00	3.00	0.0	L	.033	.031	.236	.332	.131	.076
80	1.00	0.33	0.5	L	.221	.219	.120	.019	.049	.076
80	1.00	0.50	0.5	L	.320	.319	.222	.069	.137	.189
80	1.00	1.00	0.5	L	.514	.514	.512	.352	.431	.486
80	1.00	2.00	0.5	L	.251	.247	.494	.506	.365	.291
80	1.00	3.00	0.5	L	.146	.143	.473	.576	.328	.232
80	1.00	0.33	0.7	L	.297	.292	.172	.033	.075	.109

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
80	0.67	3.00	0.0	L	.061	.027	.243	.367	.187	.124
80	0.67	0.33	0.5	L	.182	.251	.103	.024	.051	.075
80	0.67	0.50	0.5	L	.273	.340	.193	.072	.133	.173
80	0.67	1.00	0.5	L	.518	.503	.507	.383	.453	.500
80	0.67	2.00	0.5	L	.301	.218	.493	.514	.412	.345
80	0.67	3.00	0.5	L	.208	.121	.486	.588	.388	.289
80	0.67	0.33	0.7	L	.244	.321	.146	.036	.073	.104
80	0.67	0.50	0.7	L	.364	.456	.279	.103	.182	.241
80	0.67	1.00	0.7	L	.699	.712	.674	.537	.625	.679
80	0.67	2.00	0.7	L	.422	.341	.601	.622	.520	.455
80	0.67	3.00	0.7	L	.282	.179	.568	.660	.479	.370
80	0.67	0.33	0.9	L	.341	.430	.221	.064	.115	.159
80	0.67	0.50	0.9	L	.562	.640	.431	.198	.301	.371
80	0.67	1.00	0.9	L	.897	.910	.866	.755	.837	.872
80	0.67	2.00	0.9	L	.624	.541	.758	.761	.697	.651
80	0.67	3.00	0.9	L	.424	.295	.681	.757	.596	.496
80	1.50	0.33	0.0	L	.111	.081	.053	.009	.021	.034
80	1.50	0.50	0.0	L	.090	.072	.052	.010	.026	.041
80	1.50	1.00	0.0	L	.046	.055	.084	.047	.046	.048
80	1.50	2.00	0.0	L	.016	.036	.153	.173	.061	.036
80	1.50	3.00	0.0	L	.013	.030	.222	.293	.082	.039

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
80	1.50	0.33	0.5	L	.277	.224	.156	.027	.066	.104
80	1.50	0.50	0.5	L	.370	.314	.237	.067	.137	.196
80	1.50	1.00	0.5	L	.516	.512	.513	.336	.429	.485
80	1.50	2.00	0.5	L	.203	.283	.510	.495	.325	.254
80	1.50	3.00	0.5	L	.091	.161	.481	.542	.254	.164
80	1.50	0.33	0.7	L	.362	.285	.205	.039	.090	.135
80	1.50	0.50	0.7	L	.464	.405	.320	.109	.200	.266
80	1.50	1.00	0.7	L	.726	.719	.705	.513	.629	.686
80	1.50	2.00	0.7	L	.328	.415	.629	.607	.452	.381
80	1.50	3.00	0.7	L	.167	.268	.605	.639	.357	.257
80	1.50	0.33	0.9	L	.453	.380	.274	.056	.129	.182
80	1.50	0.50	0.9	L	.628	.554	.461	.173	.307	.402
80	1.50	1.00	0.9	L	.897	.885	.860	.691	.810	.860
80	1.50	2.00	0.9	L	.560	.641	.785	.754	.652	.600
80	1.50	3.00	0.9	L	.280	.408	.734	.764	.503	.383
120	1.00	0.33	0.0	L	.073	.072	.032	.002	.009	.017
120	1.00	0.50	0.0	L	.062	.061	.039	.006	.018	.028
120	1.00	1.00	0.0	L	.045	.045	.068	.045	.045	.047
120	1.00	2.00	0.0	L	.032	.032	.161	.259	.113	.067
120	1.00	3.00	0.0	L	.032	.032	.225	.449	.182	.101
120	1.00	0.33	0.5	L	.226	.224	.114	.011	.033	.055

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	1.00	0.50	0.5	L	.311	.310	.199	.044	.100	.154
120	1.00	1.00	0.5	L	.500	.500	.500	.354	.420	.465
120	1.00	2.00	0.5	L	.261	.259	.503	.592	.420	.335
120	1.00	3.00	0.5	L	.152	.150	.492	.700	.408	.274
120	1.00	0.33	0.7	L	.288	.286	.160	.017	.053	.085
120	1.00	0.50	0.7	L	.416	.415	.289	.073	.158	.220
120	1.00	1.00	0.7	L	.717	.716	.700	.513	.620	.678
120	1.00	2.00	0.7	L	.385	.382	.614	.700	.543	.449
120	1.00	3.00	0.7	L	.224	.220	.584	.764	.504	.359
120	1.00	0.33	0.9	L	.417	.414	.259	.037	.100	.136
120	1.00	0.50	0.9	L	.584	.583	.440	.128	.254	.342
120	1.00	1.00	0.9	L	.900	.900	.871	.734	.827	.869
120	1.00	2.00	0.9	L	.602	.600	.782	.816	.722	.651
120	1.00	3.00	0.9	L	.353	.347	.713	.846	.630	.493
120	0.67	0.33	0.0	L	.044	.067	.021	.002	.006	.012
120	0.67	0.50	0.0	L	.046	.061	.030	.006	.012	.020
120	0.67	1.00	0.0	L	.052	.043	.069	.052	.051	.053
120	0.67	2.00	0.0	L	.054	.030	.167	.280	.148	.091
120	0.67	3.00	0.0	L	.065	.029	.242	.468	.240	.145
120	0.67	0.33	0.5	L	.170	.231	.089	.011	.028	.048
120	0.67	0.50	0.5	L	.260	.316	.180	.043	.097	.132

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	0.67	1.00	0.5	L	.517	.517	.507	.382	.451	.492
120	0.67	2.00	0.5	L	.298	.220	.483	.595	.453	.365
120	0.67	3.00	0.5	L	.211	.133	.481	.684	.477	.344
120	0.67	0.33	0.7	L	.233	.307	.135	.018	.045	.073
120	0.67	0.50	0.7	L	.380	.454	.265	.082	.156	.217
120	0.67	1.00	0.7	L	.719	.728	.692	.548	.637	.683
120	0.67	2.00	0.7	L	.425	.340	.606	.688	.571	.483
120	0.67	3.00	0.7	L	.277	.184	.562	.751	.553	.421
120	0.67	0.33	0.9	L	.357	.446	.223	.036	.082	.126
120	0.67	0.50	0.9	L	.563	.634	.435	.146	.267	.349
120	0.67	1.00	0.9	L	.900	.912	.877	.754	.841	.875
120	0.67	2.00	0.9	L	.638	.560	.775	.820	.747	.684
120	0.67	3.00	0.9	L	.402	.289	.680	.833	.659	.532
120	1.50	0.33	0.0	L	.106	.078	.047	.003	.011	.023
120	1.50	0.50	0.0	L	.094	.074	.051	.006	.019	.034
120	1.50	1.00	0.0	L	.050	.058	.078	.047	.048	.050
120	1.50	2.00	0.0	L	.018	.038	.155	.232	.086	.044
120	1.50	3.00	0.0	L	.016	.037	.223	.414	.120	.049
120	1.50	0.33	0.5	L	.275	.217	.142	.014	.039	.071
120	1.50	0.50	0.5	L	.349	.294	.220	.047	.109	.159
120	1.50	1.00	0.5	L	.510	.510	.508	.353	.432	.478

Table 39 --continued

N	n_e/n_c	σ_e/σ_c	P	D	TT	WT	BT	CT1	CT2	JT
120	1.50	2.00	0.5	L	.204	.277	.495	.565	.361	.266
120	1.50	3.00	0.5	L	.104	.182	.504	.683	.351	.216
120	1.50	0.33	0.7	L	.355	.278	.190	.017	.053	.091
120	1.50	0.50	0.7	L	.481	.421	.324	.073	.165	.244
120	1.50	1.00	0.7	L	.714	.707	.692	.493	.609	.673
120	1.50	2.00	0.7	L	.330	.424	.628	.674	.494	.403
120	1.50	3.00	0.7	L	.168	.263	.593	.741	.426	.291
120	1.50	0.33	0.9	L	.449	.370	.272	.040	.094	.147
120	1.50	0.50	0.9	L	.631	.567	.461	.137	.271	.366
120	1.50	1.00	0.9	L	.898	.884	.867	.700	.812	.861
120	1.50	2.00	0.9	L	.558	.644	.785	.811	.699	.617
120	1.50	3.00	0.9	L	.276	.393	.727	.839	.563	.407

APPENDIX B
VARIANCE COMPONENTS FOR ESTIMATED TYPE I ERROR RATES AND POWER WHEN α=0.01 AND α=0.10

Table 40

Percent of Total Variance for Estimated Type I Error Rates Effects Accounting ≥1% of the Variance (α=0.01)

Effect	Percentage
$Mxn_e/n_c xD$	1.82
MxN	2.83
MxNxD	4.47
Mxn_e/n_c	5.60
D	6.11
M	25.18
MxD	49.97

Table 41

Percent of Total Variance for Estimated Type I Error Rates Effects Accounting ≥1% of the Variance (α=0.10)

Effect	Percentage
n_e/n_c	1.05
MxN	2.56
D	2.83
MxNxD	2.85
NxD	4.01
Mxn_e/n_c	4.23
Mxn_e/n_cxD	4.92
N	5.50
Dxn_e/n_c	7.17
M	19.55
MxD	41.87

Table 42

Percent of Total Variance for Estimated Power (Variance Differences Only) Effects Accounting ≥1% of the Variance (α=0.01)

Effect	Percentage
$n_e/n_c x \sigma_e/\sigma_c$	1.00
$N x \sigma_e/\sigma_c x D$	1.35
MxNxD	1.41
NxD	1.44
$MxNx\sigma_e/\sigma_c xD$	1.59
N	2.34
$N x \sigma_e/\sigma_c$	3.27
MxN	3.55
D	3.97
$MxNx\sigma_e/\sigma_c$	4.15
MxD	6.82
$\sigma_e/\sigma_c xD$	6.85
$M x \sigma_e/\sigma_c xD$	7.09
M	12.18
$M x \sigma_e/\sigma_c$	20.08
σ_e/σ_c	20.27

Table 43

Percent of Total Variance for Estimated Power (Variance Differences Only) Effects Accounting ≥1% of the Variance (α=0.10)

Effect	Percentage
N	1.06
MxN	1.34
MxNxσ_e/σ_c	1.75
Nxσ_e/σ_c	1.94
D	3.57
MxD	4.42
Mxσ_e/σ_cxD	5.12
σ_e/σ_cxD	7.66
M	10.05
Mxσ_e/σ_c	27.98
σ_e/σ_c	32.77

Table 44

Percent of Total Variance for Estimated Power (Mean and Variance Differences) Effects Accounting ≥1% of the Variance (α=0.01)

Effect	Percentage
D	1.04
$MxNx_e/\sigma_c$	1.34
Nx_e/σ_c	1.62
MxD	2.35
$Mx\sigma_e/\sigma_c xD$	2.45
M	2.61
$\sigma_e/\sigma_c xD$	5.31
$\sigma_e/\sigma_c xP$	5.62
P	13.18
$Mx\sigma_e/\sigma_c$	21.48
σ_e/σ_c	37.88

Table 45

Percent of Total Variance for Estimated Power (Mean and Variance Differences) Effects Accounting ≥1% of the Variance (α=0.10)

Effect	Percentage
M x σ_e/σ_c x D	2.45
σ_e/σ_c x D	5.31
P	10.80
M x σ_e/σ_c	23.15
σ_e/σ_c	54.85

REFERENCES

Algina, J.J., Oshima, T.C., and Lin, W.Y. (1994). Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. Journal of Educational and Behavioral Statistics, 19, 275-291.

Berry, K.J. and Mielke, P.W. (1992). A family of multivariate measures of association for nominal independent variables. Educational and Psychological Measurement, 52, 41-55.

Best, D.J and Rayner, J.C.W. (1987). Welch's approximate solution for the Behrens-Fisher problem. Technometrics, 29, 205-210.

Blair, R.C., and Sawilowsky, S. (1993). A note on the operating characteristics of the modified F test. Biometrics, 49, 935- 939.

Boos, D.D. and Brownie, C. (1986). Testing for a treatment effect in the presence of nonresponders. Biometrics, 42, 191-197.

Bradley, J.V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Brownie, C., Boos, D.D., and Hughes-Oliver, J. (1990). Modifying the t and ANOVA F tests when treatment is expected to increase variability relative to controls. Biometrics, 46, 259-266.

Chirembo, A.M. (1995). Direct versus indirect methods for the estimation of variance-covariance matrices and regression parameters when data are skewed and incomplete. Unpublished doctoral dissertation, University of Florida.

Conover, W.J., and Salsburg, D.S. (1988). Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to "respond" to treatment. Biometrics, 44, 198-196.

Coombs. W.T. (1992) Solutions to the multivariate G-sample Berens-Fisher problem based upon generalizations of

the Brown-Forsythe F* and Wilcox H_a tests. Dissertation Abstaracts International, 54(1), 313B.

Efron, B. (1969). Student's t-test under symmetry conditions. Journal of the American Statistical Association, 64, 1278-1308.

Fisher, R.A, (1939). The comparison of samples with possibly unequal variances. The Annals of Eugenics, 9, 174-180.

Good, P.I. (1979). Detection of treatment effect when not all experimental subjects will respond to treatment. Biometrics, 35, 483-489.

Hodges, J.L., Jr and Lehmann, E.L (1956). The efficiency of some nonparametric competitors of the t-test. Annals of Mathematical Statistics, 27, 234-335.

Johnson, R.A., Verrill, S., and Moore, D.H. II (1987). Two-sample rank tests for detecting changes that occur in a small proportion of treated population. Biometrics, 43, 641-655.

Kendall, M.G., and Stuart, A. (1973). The advanced theory of statistics (Volume 2). New York: Hafner.

Lehmann, E.L. (1953). The power of rank tests. Annals of Mathematical Statistics, 24, 23-43.

Lehmann, E.L. (1959). Testing statistical hypotheses. John Wiley and Sons, Inc. New York.

Miceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Mielke, P. W. (1991). The application of multivariate permutation methods based on distance functions in the earth sciences. Earth-Science Reviews, 13, 377-387.

Mielke, P.W., and Berry, K.J. (1994). Permutation tests for common locations among samples with unequal variances. Journal of Educational and Behavioral Statistics, 3, 217-236.

Mood, A.M., Graybill, F.A., & Boes, D.S. (1974). Introduction to the theory of statistics. New-York: McGraw-Hill.

Myers, J.L. (1979). Fundamentals of experimental design. Boston: Allyn and Bacon.

Salsburg, D.S. (1986). Alternative hypothesis for effects of drugs in small-scale clinical studies. Biometrics, 42, 671- 674.

SAS Institute Inc. (1985). SAS user's guide: Basics, Version 5 edition. Cary, NC: Author.

Subrahmaniam, K., Subrahmaniam, K. and Messeri, J.Y. (1975). On the robustness of some tests of significance in sampling from compound normal population. Journal of the American Statistical Association, 70, 435-438.

BIOGRAPHICAL SKETCH

Mbulaheni E. Nthangeni was born in Sibasa (Venda), South Africa, on May 04, 1954. He received his B.Sc. and B.Sc. Hons. from the University of the North, South Africa, in 1980 and 1982, respectively. In 1987, he received his M.Sc. in operations research from the University of South Africa, South Africa. From 1982-1988 he lectured in the Department of Mathematical Statistics at the University of the North, South Africa.

In Fall 1988, he enrolled for graduate studies in the Department of Industrial and Systems Engineering at the University of Florida, USA. He received an M.S. in industrial and systems engineering, in 1991. In Spring 1992, he began studying for the Ph.D. in research and evaluation methodology in the Department of Foundations of Education, University of Florida. He will graduate in December, 1995.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James J. Algina, Chair
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Linda M. Crocker
Professor of Foundations
of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

M. David Miller
Associate Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sencer Yeralan
Associate Professor of
Industrial and Systems
Engineering

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1995

Chair, Foundations of Education

Dean, College of Education

Dean, Graduate School